Convex Stochastic optimization WiSe 2024/2025

Ari-Pekka Perkkiö

January 14, 2025

Contents

1 Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t=0}^T$ of sub-σ-algebras of $\mathcal{F}$ and consider the dynamic stochastic optimization problem

$$\text{minimize} \quad Eh(x) := \int h(x(\omega), \omega) dP(\omega) \quad \text{over } x \in \mathcal{N}, \tag{SP}$$

where, for given integers n_t and m,

$$\mathcal{N} = \{(x_t)_{t=0}^T \mid x_t \in L^0(\Omega, \mathcal{F}_t, P; \mathbb{R}^{n_t})\}$$

is the space of *adapted processes*, h is an extended real-valued $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable function, where $n := n_0 + \ldots + n_T$. More precisely, h is a normal integrand that will be defined later on. Here and in what follows, we define the expectation of a measurable function ϕ as $+\infty$ unless the positive part ϕ^+ is integrable[1]. The function Eh is thus well-defined extended real-valued function on $\mathcal{N}$.

We will assume throughout that the function $h(\cdot, \omega)$ is *convex* for every $\omega \in \Omega$. It will turn out that Eh is a convex function of $\mathcal{N}$ and (SP) is a *convex stochastic optimization problem* on the space of adapted processes.

The aim of this course is to analyze (SP) using *dynamic programming* and *conjugate duality*. These lead to characterizations of optimal solutions of (SP) via "Bellman equations" and "Karush-Kuhn-Tucker conditions", both starting points of various modern numerical methods. Duality theory also leads to characterizations and lower bounds of the optimal value of (SP), a classical example being that the superhedging price of an option in a liquid market can be computed via "martingale measures".

1.1 Examples

Throughout the course, we will be returning to the examples that are presented in this section. Undefined concepts and nontrivial claims will be given a rigorous treatment later on.

Example 1.1 (Mathematical programming)**.** *Consider the problem*

$$\begin{array}{ll} \text{minimize} & Ef_0(x) \quad \text{over } x \in \mathcal{N} \\ \text{subject to} & f_j(x) \le 0 \;\; P\text{-}a.s., \;\; j = 1, \ldots, m, \end{array}$$

where f_j are normal integrands. The problem fits the general framework with

$$h(x, \omega) = \begin{cases} f_0(x, \omega) & \text{if } f_j(x, \omega) \le 0 \text{ for } j = 1, \ldots, m, \\ +\infty & \text{otherwise.} \end{cases}$$

[1] In particular, the sum of extended real numbers is defined as $+\infty$ if any of the terms equals $+\infty$.

When each $f_0(\cdot,\omega),\dots,f_m(\cdot,\omega)$ *is an affine function, the problem becomes* a linear stochastic optimization problem.

Given a stochastic process x, $\Delta x_t := x_t - x_{t-1}$ is the *backward difference* at time t.

Example 1.2 (Optimal stopping)**.** *Consider the problem*

$$\underset{x\in\mathcal{N}_+}{\text{maximize}} \quad E\sum_{t=0}^{T} Z_t\Delta x_t \quad \text{subject to} \quad \Delta x \geq 0,\ x \leq 1\ P\text{-}a.s.$$

for an adapted real-valued process Z *and* $x_{-1} := 0$*, This is a convex relaxation of the* optimal stopping problem

$$\underset{\tau\in\mathcal{T}}{\text{maximize}} \quad EZ_\tau,$$

where $\mathcal{T}$ *is the set of stopping times. This fits the general framework with,* $n_t = 1$ *for all* t *and*

$$h(x,\omega) = \begin{cases} -\sum_{t=0}^{T} Z_t(\omega)\Delta x_t & \textit{if } \Delta x \geq 0 \textit{ and } x \leq 1,\\ +\infty & \textit{otherwise,}\end{cases}$$

Example 1.3 (Optimal investment)**.** *Let* $s = (s_t)_{t=0}^T$ *be an adapted* $\mathbb{R}^J$*-valued stochastic process describing the unit prices or assets in a perfectly liquid financial market. Consider the problem of finding a dynamic trading strategy* $x = (x_t)_{t=0}^T$ *that provides the "best hedge" against a financial liability of delivering a random amount* $c \in L^0$ *cash at time* T*. If we measure our risk preferences over random cash-flows with the "expected shortfall" associated with a nondecreasing convex "loss function"* $V : \mathbb{R} \to \overline{\mathbb{R}}$*, the problem can be written as*

$$\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}_D, \tag{1.1}$$

where $\mathcal{N}_D$ *denotes the set of adapted trading strategies* $x = (x_t)_{t=0}^T$ *that satisfy the portfolio constraints* $x \in D_t$ *for all* $t = 0,\dots,T$ *almost surely. Here* D_t *is a random* $\mathcal{F}_t$*-measurable set consisting of the portfolios we are allowed to hold over time period* $(t, t+1]$*.*

The problem fits the general framework with

$$h(x,\omega) = \begin{cases} V\left(u(\omega) - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}(\omega)\right) & \textit{if } x_t \in D_t(\omega) \textit{ for } t = 0,\dots,T\\ +\infty & \textit{otherwise.}\end{cases}$$

This example can be extended to a semi-static hedging problem*, where some (or all) of the assets are allowed to be traded only at the initial time* $t = 0$*. It is also possible to allow some of the assets to be "American type options". The above could also be readily extended by allowing the loss function* V *to be random or by adding transaction costs.*

Much of financial mathematics has revolved around the problem of assigning values to financial products that provide a random payout $c \in L^0$ at a future date T. A classical approach is "superhedging".

Example 1.4 (Superhedging)**.** *Consider the problem*

$$\begin{array}{ll} \text{minimize} & \alpha \quad \text{over } \alpha \in \mathbb{R},\ x \in \mathcal{N} \\ \text{subject to} & \alpha + \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \geq c \quad a.s., \\ & x_t \in D_t \quad t = 0, \ldots, T \ a.s. \end{array} \tag{1.2}$$

This is the classical superhedging problem of finding the least amount of initial capital α that can finance a self-financing trading strategy x whose liquidation value at time T exceeds the liability c almost surely.

This is an instance of (**??**) *but with time t running from -1 to T, $\mathcal{F}_{-1} = \{\Omega, \emptyset\}$, $x_{-1} = \alpha$ and*

$$h(\alpha, x, \omega) = \alpha + \sum_{t=0}^{T} \delta_{D_t(\omega)}(x_t) + \delta_{\mathbb{R}_+}(\alpha + \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega) - c(\omega)).$$

Classical "risk-neutral" valuations in financial mathematics can be seen as a special case of superhedging. We will see that is is an application of duality theory so that the superhedging cost can be expressed in terms of expectations of c under so called "equivalent martingale measures". In practice, however, the requirement of superhedging is often unreasonable and the associated cost is too high to be competitive. More practical approach is to use indifference pricing.

Remark 1.5 (Indifference pricing)**.** *The* indifference selling price *of a claim $c \in L^0$ is defined by*

$$\pi(\bar{c}; c) := \inf\{\alpha \in \mathbb{R} \mid \varphi(\bar{c} + c - \alpha) \leq \varphi(\bar{c})\},$$

where $\varphi(c)$ denotes the optimum value of (3.4)*. Here $\bar{c} \in L^0$ denotes the traders initial liability cashflows and α is the price she would receive in compensation of delivering an additional random cashflow $c \in L^0$ (a "contingent claim"). The indifference selling price $\pi(\bar{c}; c)$ is the least price at which it would make sense for the trader to sell the claim c for. The indifference buying price is defined analogously. If $\bar{c} = 0$, $V = \delta_{\mathbb{R}_-}$ and $\varphi(0) = 0$, the indifference price becomes the superhedging cost. When $V = \delta_{\mathbb{R}_-}$, the condition $\varphi(0) = 0$ means that one cannot turn a strictly negative initial wealth into a random terminal wealth that is nonnegative almost surely.*

The indifference pricing principle makes good sense also in more general market models. The following extends problem (3.4) by allowing for investments in a finite set of contingent claims that can be traded at time $t = 0$ at a cost given by a convex function S_0.

Example 1.6 (Semi-static hedging)**.** *Consider the problem*

$$\begin{aligned}&\text{minimize} \quad EV\left(c-\sum_{t=0}^{T-1}x_t\cdot\Delta s_{t+1}-\bar c\cdot\bar x+S_0(\bar x)\right) \text{ over } x\in\mathcal{N},\ \bar x\in\mathbb{R}^{\bar J},\\&\text{subject to} \quad x_t\in D_t \quad t=0,\dots,T \ a.s.,\end{aligned} \tag{1.3}$$

where s, D and $\mathcal{N}$ are as in (3.4) and $\bar J$ is another finite set of assets that are traded only at time $t=0$. The portfolio $\bar x\in\mathbb{R}^{\bar J}$ is bought before the dynamic trading of the assets J starts and it is held fixed (static) until time T. The random vector $\bar c$ gives the payouts of the statically held contingent claims and the function $S_0:\mathbb{R}^{\bar J}\to\overline{\mathbb{R}}$ gives the cost of buying the portfolio $\bar x$ at the best available market prices. We assume that S_0 is a proper lsc convex function that vanishes at the origin.

*Problem (1.3) fits the format of (**??**) with time t running from -1 to T, $\mathcal{F}_{-1}=\{\Omega,\emptyset\}$, $x_{-1}=\bar x$ and*

$$h(\bar x,x,\omega)=V\left(c(\omega)-\sum_{t=0}^{T-1}x_t\cdot\Delta s_{t+1}(\omega)-\bar c(\omega)\cdot\bar x+S_0(\bar x),\omega\right)+\sum_{t=0}^{T-1}\delta_{D_t(\omega)}(x_t,\omega).$$

Convexity of S_0 arises naturally in practice. For example, if the buying and selling prices of the claims $\bar c$ are given by vectors $s^b\in\mathbb{R}^{\bar J}$ and $s^a\in\mathbb{R}^{\bar J}$, respectively, and if we assume that one can buy and sell infinite quantities at these prices, then

$$S_0(x)=\sup_{s\in[s^b,s^a]}x\cdot s.$$

If the bid and ask prices come with finite quantities given by vectors $q^b\in\mathbb{R}^{\bar J}$ and $q^a\in\mathbb{R}^{\bar J}$, respectively, then

$$S_0(x)=\sup_{s\in[s^b,s^a]}x\cdot s+\delta_{[-q^b,q^a]}(x).$$

More generally, the cost of buying a portfolio $\bar x$ in limit order markets always results in a proper lsc convex cost function S_0.

Example 1.7 (Stochastic control)**.** *The problem*

$$\begin{aligned}&\text{minimize} \quad E\left[\sum_{t=0}^{T}L_t(X_t,U_t)\right] \quad \text{over } X,U\in\mathcal{N},\\&\text{subject to} \quad \Delta X_t=A_tX_{t-1}+B_tU_{t-1}+u_t \quad t=1,\dots,T\end{aligned}$$

fits the general framework with $x=(X,U)$,

$$h(x,\omega)=\begin{cases}\sum_{t=0}^{T}L_t(x_t,\omega) & \text{if } \pi\Delta x_t-\bar A_t(\omega)x_{t-1}=u_t(\omega) \text{ for } t=1,\dots,T,\\+\infty & \text{otherwise,}\end{cases}$$

where $\pi = [I\ 0]$ and $\bar{A}_t = [A_t\ B_t]$. Here the stochastic process X is the "state" and U is the "control".

Example 1.8 (Problems of Lagrange)**.** *Consider the problem*

$$\underset{x \in \mathcal{N}}{\text{minimize}} \quad E \sum_{t=0}^{T} K_t(x_t, \Delta x_t), \tag{1.4}$$

where $n_t = d$, $\Delta x_t := x_t - x_{t-1}$, $x_{-1} := 0$.

The problem fits the general framework with

$$h(x, \omega) = \sum_{t=0}^{T} K_t(x_t, \Delta x_t, \omega).$$

For instance, currency market models *fit into this framework, where components of x_t describe different currencies in the portfolio hold at time $(t, t+1]$.*

Example 1.9 (Electricity storage management)**.** *Consider the problem*

$$\begin{aligned} &\text{maximize} \quad && Eu(X_T^m) \quad \text{over } (X, U) \in \mathcal{N} \\ &\text{subject to} && X_0^m = x_0, \\ &&& X_0^e = 0, \\ &&& X_{t+1}^m \le R_{t+1}^m(X_t^m) - S_{t+1}(U_t^e), \\ &&& X_{t+1}^e \le R_{t+1}^e(X_t^e, U_t^e), \end{aligned} \tag{1.5}$$

where

- *X_t^e is the amount of energy in storage at time t,*
- *X_t^m is the amount of money market investments at time t, $x_0 \in \mathbb{R}$ is the initial wealth.*
- *U_t^e is the amount of energy bought at time t*

The function u is a utility function, i.e., an extended-real valued concave increasing function. The function S_t determines how much it costs to buy energy at time t and it depends on the state of the electricity market. A negative value of U_t^e means that $-U_t^e$ units of energy is withdrawn from the storage in order to sell electricity in the market. The cost $S_t(U_t^e)$ would then be negative which means that the agent generates revenue that can be invested in the money market. The function R_t^m describes the interest structure of the moeny market while functions R_t^e determines how many electricity is stored given the current storage level and the electrity bought. The function R_t^e depends on the physical properties of the storage (very different for a lithium battery, hydrogen storage or a pump station). Negative investment in the money market is interpreted as borrowing as usual.

The functions S_t, R^e_t and R^m_t are random, in general. In particular, S_t depends on prevailing electricity prices while R^m_t depends on the available money market rates both of which are typically modeled by stochastic processes. Under appropriate assumption on the functions R^m, R^e and S, the problem can be written as a stochastic control problem.

Example 1.10 (Risk measures)**.** *Some optimization problems in finance are given in terms of risk measures that do not a priori fit into (**??**). However, some of them can be expressed as (**??**) by introducing additional variables.*

Consider the problem

$$\underset{x\in\mathcal{N}}{\text{minimize}} \quad \mathcal{V}(x) \tag{1.6}$$

for $\mathcal{V} : L^0 \to \overline{\mathbb{R}}$ defined by

$$\mathcal{V}(x) = \inf_{\alpha\in\mathbb{R}} Ec(\alpha, x),$$

where $c(\cdot, \cdot, \omega)$ is convex on $\mathbb{R} \times \mathbb{R}^n$.

Extending $\bar{\mathcal{N}} := L^0(\mathcal{F}_{-1}) \times \mathcal{N}$ for a trivial $\mathcal{F}_{-1}$ and denoting $\bar{x} = (\alpha, x)$, the problem fits the general framework with

$$h(\bar{x}, \omega) = c(\alpha, x, \omega).$$

Assume now that $c(\alpha, x, \omega) = \alpha + \theta(g(x, \omega) - \alpha, \omega)$ for convex θ and g. When θ is nondecreasing with $\theta(0) = 0$ and $1 \in \partial\theta(0)$,

$$\mathcal{V}(x) = \inf_{\alpha} E[\alpha + \theta(g(x) - \alpha)]$$

is known as optimized certainty equivalent *of the random variable $g(x)$. When $\theta(u) = e^t - 1$, we get the* entropic risk

$$\mathcal{V}(x) = \log Ee^{g(x)}$$

while for $\theta(u) = \frac{u^+}{\gamma}$ with $\gamma \in (0, 1)$ we obtain the conditional value at risk at γ

$$\mathcal{V}(x) = \inf_{\alpha} E[\alpha + \frac{1}{\gamma}(g(x) - \alpha)^+].$$

When g is affine and $\theta(u) = \frac{1}{2}u^2 + u$, we obtain the ”Mean Variance” risk measure

$$\mathcal{V}(x) = \frac{1}{2}E[(g(x) - Eg(x))^2] + Eg(x).$$

Exercise 1.1.1. *Verify that one really gets entropic risk, conditional value at risk and mean variance in the above example.*

2 Convex analysis

2.1 Convex sets

Let U be a real vector space. A set $A \subseteq U$ is *convex* if

$$\lambda u + (1-\lambda)u' \in A.$$

for every $u, u' \in A$ and $\lambda \in (0,1)$. For $\lambda \in \mathbb{R}$ and sets A and B, we define the scalar multiplication and a sum of sets as

$$\begin{aligned} \lambda A &:= \{\lambda u \mid u \in A\} \\ A + B &:= \{u + u' \mid u \in A,\, u' \in B\}. \end{aligned}$$

The summation is also known as *Minkowski addition*. With this notation, A is convex if and only if

$$\lambda A + (1-\lambda)A \subseteq A \quad \forall \lambda \in (0,1).$$

Convex sets are stable under many algebraic operations. Let X be another linear vector space.

Theorem 2.1. *Let $\mathcal{J}$ be an arbitrary index set, $(A^j)_{j\in\mathcal{J}}$ a collection of convex sets and $A \subset X \times U$ a convex set. Then,*

1. *for $\lambda \in \mathbb{R}_+$, the scaled set λA is convex,*
2. *for finite $\mathcal{J}$, the sum $\sum_{j\in\mathcal{J}} A^j$ is convex,*
3. *the intersection $\bigcap_{j\in\mathcal{J}} A^j$ is convex,*
4. *the projection $\{u \in U \mid \exists x : (x,u) \in A\}$ is convex.*

Proof. Exercise. □

2.2 Locally convex topological vector spaces

Next we turn to topological properties. Let τ be a topology on U (the collection of *open sets*, their complements are called *closed sets*) and let $A \subset U$. The *interior* $\operatorname{int} A$ of A is the union of all open sets contained in A and *closure* $\operatorname{cl} A$ is the intersection of closed sets containing A.

The set A is a *neighborhood* of u if $u \in \operatorname{int} A$. We denote the collection of neighborhoods of u by $\mathcal{H}_u$ and the collection of open neighborhoods of u by $\mathcal{H}_u^o$. Note that A is a neighborhood of u if and only if A contains an open neighborhood of u.

Exercise 2.2.1. *For $A \subset U$, $u \in \operatorname{cl} A$ if and only if $A \cap O \neq \emptyset$ for all $O \in \mathcal{H}_u^o$.*

A function g from U to another topological space V is *continuous at a point* u if the preimage of every neighborhood of $g(u)$ is a neighborhood of u. A function f is *continuous* if it is continuous at every point.

Exercise 2.2.2. *A function is continuous if and only if the preimage of every open set is open.*

A collection $\mathcal{E}$ of neighborhoods of u is called a *neighborhood base* if every neighborhood of u contains and element of $\mathcal{E}$. Evidently $\mathcal{H}_u^0$ is a neighborhood base.

Exercise 2.2.3. *Given local bases $\mathcal{E}_u$ of u and $\mathcal{E}_v$ of $v = g(u)$, g is continuous at u if and only if the preimage of every element of $\mathcal{E}_v$ contains an element of $\mathcal{E}_u$.*

Given another topological space (U', τ'), the *product topology* on $U \times U'$ is the smallest topology containing all the sets $\{O \times O' \mid O \in \tau, O' \in \tau'\}$. We always equip products of topological spaces with the product topology. Clearly $\{(O, O') \in \mathcal{H}_u^o \times \mathcal{H}_{u'}^o\}$ is a neighborhood basis of (u, u').

Exercise 2.2.4. *Let $p : U \times U' \to V$ be continuous. For every $u' \in U'$, $u \mapsto p(u, u')$ is continuous.*

The space (U, τ) is a *topological vector space* (TVS) if $(u, u') \to u + u'$ is continuous from $U \times U$ to U and $(u, \alpha) \to \alpha u$ is continuous from $U \times \mathbb{R}$ to U.

Exercise 2.2.5. *In a topological vector space U,*

1. *$\alpha O \in \mathcal{H}_0^0$ for all $\alpha \neq 0$ and $O \in \mathcal{H}_0^o$,*
2. *for all $u \in U$ and $O \subset U$, $(O + u) \in \mathcal{H}_u^o$ if and only if $O \in \mathcal{H}_0^o$,*
3. *sum of a nonempty open set with any set is open,*
4. *for every $O \in \mathcal{H}_0^o$, there exists $O' \in \mathcal{H}_0^o$ such that $2O' \subset O$,*
5. *$\alpha A \in \mathcal{H}_0$ for all $\alpha \neq 0$ and $A \in \mathcal{H}_0$,*
6. *for all $u \in U$ and $A \in \mathcal{U}$, $(A + u) \in \mathcal{H}_u$ if and only if $A \in \mathcal{H}_0$,*
7. *for every $A \in \mathcal{H}_0$, there exists $A' \in \mathcal{H}_0$ such that $2A' \subset A$.*
8. *for every $\lambda \in (0, 1)$ and $O \in \mathcal{H}_0^0$, there exists $O^1, O^2 \in \mathcal{H}_0^o$ such that $\lambda O^1 + (1 - \lambda) O^2 \subset O$.*

A set C is *symmetric* if $x \in C$ implies $-x \in C$.

Lemma 2.2. *In a topological vector space, every (resp. convex) neighborhood of the origin contains a symmetric (resp. convex) neighborhood of the origin.*

Proof. Let $A \in \mathcal{H}_0$. By continuity of $p(\alpha, u) := \alpha u$ from $\mathbb{R} \times U$ to U, there is α' and $O \in \mathcal{H}_0^o$ such that $\alpha O \subset A$ for all $|\alpha| \le \alpha'$. The set $B := \bigcup_{|\alpha| \le \alpha'} (\alpha O)$ is the sought neighborhood.

Assume additionally that A is convex. The set $A \cap (-A)$ is symmetric, so, since $B \subset A$ is symmetric as well, $B \subset A \cap (-A)$. Hence $A \cap (-A)$ is a symmetric convex set containing a neighborhood of the origin. □

Lemma 2.3. *Let C be a convex set in a TVS. Then* $\operatorname{int} C$ *and* $\operatorname{cl} C$ *are convex.*

Proof. Let $\lambda \in (0,1)$. We have $\operatorname{int} C \subset C$, so $\lambda(\operatorname{int} C) + (1-\lambda) \operatorname{int} C \subset C$. Since sums and strictly positive scalings of open sets are open, we see that $\lambda(\operatorname{int} C) + (1-\lambda) \operatorname{int} C \subset \operatorname{int} C$, since $\operatorname{int} C$ is the largest open set contained in C. Since $\lambda \in (0,1)$ was arbitrary, this means that $\operatorname{int} C$ is convex.

To prove that $\operatorname{cl} C$ is closed we use results from Exercises 2.2.1 and 2.2.5. Let $u, u' \in \operatorname{cl} C$, $\lambda \in (0,1)$ and $\tilde{O} \in \mathcal{H}_0^o$. It suffices to show that

$$\lambda u + (1-\lambda)u' + \tilde{O} \cap C \neq \emptyset.$$

There are $O, O' \in \mathcal{H}_0^o$ with $\lambda O + (1-\lambda)O' \subset \tilde{O}$ and $\tilde{u} \in C \cap (u + O)$ and $\tilde{u}' \in C \cap (u' + O')$. Thus

$$\lambda \tilde{u} + (1-\lambda)\tilde{u}' \subset \lambda(u + O) + (1-\lambda)(u' + O') \subset \lambda u + (1-\lambda)u' + \tilde{O}$$

where the left side belongs to C. □

2.3 Separation theorems

Sets of of the form

$$\{u \in U \mid l(u) = \alpha\}$$

are called hyper-planes, where l is a real-valued linear function and $\alpha \in R$. Each hyperplane generates two *half-spaces* (opposite sides of the plane)

$$\{u \in U \mid l(u) \le \alpha\}, \quad \{u \in U \mid l(u) \ge \alpha\}.$$

A hyperplane *separates* sets C^1 and C^2 if they belong to the opposite sides of the hyperplane. The separation is *proper* unless both sets are contained in the hyperplane. In other words, proper separation means that

$$\sup\{l(u^1 - u^2) \mid u^i \in C^i\} \le 0 \quad \text{and} \quad \inf\{l(u^1 - u^2) \mid u^i \in C^i\} < 0.$$

A set $C \subset U$ is called *algebraically open* if $\{\alpha \in \mathbb{R} \mid u + \alpha u' \in C\}$ is open for any $u, u' \in U$. The set C is *algebraically closed* if its complement is open, or equivalently, if the set $\{\alpha \in \mathbb{R} \mid u + \alpha u' \in C\}$ is closed for any $u, u' \in U$.

Exercise 2.3.1. *In a topological vector space, open (resp. closed) sets are algebraically open (resp. closed), and the sum of a nonempty algebraically open set with any set is algebraically open.*

The following separation theorem states that the origin and an algebraically open convex set not containing the origin and can be properly separated.

Theorem 2.4. *Assume that C in a linear vector space U is an algebraically open convex set with $0 \notin C$. Then there exists a linear $l : U \to \mathbb{R}$ such that*

$$\sup\{l(u) \mid u \in C\} \leq 0, \quad \inf\{l(u) \mid u \in C\} < 0.$$

In particular, $l(u) < 0$ for all $u \in C$.

Proof. This is an application of Zorn's lemma. Omitted. □

The above separation theorem implies a series of other separation theorems for convex sets. In the locally convex setting below, we get separation theorems in terms of continuous linear functionals, or equivalently, in terms of closed hyperspaces as the next exercise shows.

A real-valued function g is *bounded from above* on $B \subset U$ if there is $M \in \mathbb{R}$ such that $g(u) < M$ for all $u \in B$. If g is continuous at $u \in \operatorname{dom} g$, then it is bounded from above on a neighborhood at u. Indeed, choose a neighborhood $g^{-1}((-\infty, M))$ for some $M > g(u)$.

Theorem 2.5. *Assume that l is a real-valued linear function on a topological vector space. Then the following are equivalent:*

1. *l is bounded from above in a neighborhood of the origin.*
2. *l is continuous.*
3. *$\{u \in U \mid l(u) = \alpha\}$ is closed for all $\alpha \in \mathbb{R}$.*
4. *$\{u \in U \mid l(u) = 0\}$ is closed.*

Proof. Exercise. □

A topological vector space is *locally convex* (LCTVS) if every neighborhood of the origin contains a convex neighborhood of the origin.

Theorem 2.6. *Assume that C is a closed convex set in a LCTVS and $u \notin C$. Then there is a continuous linear functional separating properly u and C.*

Proof. The origin belongs to the open set $(C - u)^C$, so there is a convex $O \in \mathcal{H}_0^o$ such that $0 \notin C - u + O$. By Theorem 2.4, there is a linear l such that

$$l(u') < 0 \quad \forall\, u' \in C - u + O.$$

This means that $l(u') < l(u)$ for all $u' \in C + O$, so l is continuous by Theorem 2.5. □

The following corollary is very important in the sequel. For instance, it will give the biconjugate theorem that is the basis of duality theory in convex optimization.

Corollary 2.7. *The closure of convex set in a LCTVS is the intersection of all closed hyperplanes containing the set.*

Proof. By Lemma 2.3, $\operatorname{cl} C$ is convex for convex C. For any $u \notin \operatorname{cl} C$, there is, by the above theorem, a closed half-space H_u such that $\operatorname{cl} C \subset H_u$ and $u \notin H_u$. We get

$$\operatorname{cl} C = \bigcap_{u \notin C} H_u.$$

□

2.4 Convex functions

Throughout the course, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is the *extended real line.* For $a, b \in \overline{\mathbb{R}}$, the ordinary summation is extended as $a + b = +\infty$ if $a = +\infty$, and as $a + b = -\infty$ if $a \neq +\infty$ and $b = -\infty$.

Let $g : U \to \overline{\mathbb{R}}$. The function g is *convex* if

$$g(\lambda u + (1 - \lambda)u') \leq \lambda g(u) + (1 - \lambda)g(u')$$

for all $u, u' \in U$ and $\lambda \in [0, 1]$. A function is convex if and only if its *epigraph*

$$\operatorname{epi} g := \{(u, \alpha) \in U \times \mathbb{R} \mid g(u) \leq \alpha\}$$

is a convex set. Applying the last part of Theorem 2.1 to $\operatorname{epi} g$, we see that the *domain*

$$\operatorname{dom} g := \{u \in U \mid g(u) < \infty\}$$

is convex when g is convex.

Exercise 2.4.1. *Given $g : U \to \overline{\mathbb{R}}$, the following are equivalent:*

1. *g is convex;*
2. $\operatorname{epi} g$ *is convex;*
3. *the* strict epigraph

   $$\operatorname{epi}_s g = \{(u, \alpha) \in \mathcal{U} \times \mathbb{R} \mid g(u) < \alpha\}$$

   is convex.

Many algebraic operations also preserve convexity of functions.

Theorem 2.8. *Let $\mathcal{J}$ be an arbitrary index set, $(g^j)_{j\in\mathcal{J}}$ a collection of convex functions, and $p : X \times U \to \overline{\mathbb{R}}$ a convex function. Then,*

1. *for finite $\mathcal{J}$ and strictly positive $(\lambda^j)_{j\in\mathcal{J}}$, the sum $\sum_{j\in\mathcal{J}} \lambda^j g^j$ is convex,*
2. *the* infimal convolution

$$u \mapsto \inf\{\sum g^j(u^j) \mid \sum u^j = u\}$$

   is convex,

3. *the supremum $u \mapsto \sup_{j\in\mathcal{J}} g^j(u)$ is convex,*
4. *the* marginal function $u \mapsto \inf_x p(x,u)$ *is convex.*

Proof. Exercise. □

The function g is called *positively homogeneous* if

$$g(\alpha u) = \alpha g(u) \quad \forall\ u \in \operatorname{dom} g \text{ and } \forall\ \alpha > 0,$$

and *sublinear* if

$$g(\alpha^1 u^1 + \alpha^2 u^2) \leq \alpha^1 g(u^1) + \alpha^2 g(u^2) \qquad \forall\ u^i \in \operatorname{dom} g \text{ and } \forall\ \alpha^i > 0.$$

The second part in the following exercise shows that norms are convex.

Exercise 2.4.2. *Let g be an extended real-valued function on U.*

1. *If g is positively homogeneous and convex, then it is sublinear.*
2. *If g is positively homogeneous, then it is convex if and only if*

$$g(u^1 + u^2) \leq g(u^1) + g(u^2) \qquad \forall\ u^i \in \operatorname{dom} g.$$

3. *If g is convex, then*

$$G(\lambda, u) = \begin{cases} \lambda g(u/\lambda) & \text{if } \lambda > 0, \\ +\infty & \text{otherwise} \end{cases}$$

   is positively homogeneous and convex on $\mathbb{R} \times U$. In particular,

$$p(u) = \inf_{\lambda > 0} G(\lambda, u)$$

   is positively homogeneous and convex on U.

The third part above is sometimes a surprising source of convexity. It also implies properties for recession functions and directional derivatives introduced later on.

Let G be a function from a subset $\operatorname{dom} G$ of X to U and let $K \subset U$ be a convex cone. The function G is *K-convex* if

$$\operatorname*{epi}_K G := \{(x, u) \mid x \in \operatorname{dom} G, G(x) - u \in K\}$$

is a convex set in $X \times U$. Note that $\operatorname{dom} G$ is convex, being the projection of $\operatorname{epi}_K G$ to X. When $G : X \to \mathbb{R}$, G is convex if and only if it is K-convex for $K = \mathbb{R}_-$, in which case $\operatorname{epi}_K G = \operatorname{epi} G$.

Lemma 2.9. *The function G is K-convex if and only if $\operatorname{dom} G$ is convex and*

$$G(\lambda x_1 + (1-\lambda)x_2) - \lambda G(x_1) - (1-\lambda)G(x_2) \in K$$

for every $x_i \in \operatorname{dom} G$ and $\lambda \in (0, 1)$

Proof. Exercise. □

The *composition* $g \circ G$ of g and G is defined by

$$\operatorname{dom}(g \circ G) := \{x \in \operatorname{dom} G \mid G(x) \in \operatorname{dom} g\}$$
$$(g \circ G)(x) := g(G(x)) \quad \forall x \in \operatorname{dom}(g \circ G).$$

The range of G is denoted by $\operatorname{rge} G$.

Theorem 2.10. *If G is K-convex and g is convex such that $g(u_1) \leq g(u_2)$ whenever $u_1 \in \operatorname{rge} G$ and $u_1 - u_2 \in K$, then $g \circ G$ is convex.*

Proof. Exercise. □

Exercise 2.4.3. *Let G a convex function on X and h a nondecreasing convex function on $\operatorname{rge} G$. Then $h \circ G$ is convex.*

Exercise 2.4.4. *The function $g(u) = \Pi_{i=1}^n u_i^{\lambda_i}$ is concave on $\mathbb{R}^n$, where $u = (u_1, \ldots, u_n)$, $\lambda_i > 0$ and $\sum_{i=1}^n \lambda_i < 1$.*

Hint: When $\sum_{i=1}^n \lambda_i = 1$, apply Exercise 2.4.2. For the general case, combine with the composition rule.

2.5 Lower semicontinuity, recessions and directional derivatives

Assume now that g is an extended real-valued function on U. The function g is said to be *proper* if it is not identically $+\infty$ and if it never takes the value $-\infty$. The function g is *lower semicontinuous* (lsc) if the *level-set*

$$\operatorname{lev}_\alpha g := \{u \in U \mid g(u) \leq \alpha\}$$

is closed for each $\alpha \in \mathbb{R}$. Equivalently, g is lsc if its epigraph is closed, or if, for every $u \in U$,

$$\sup_{A \in \mathcal{H}_u} \inf_{u' \in A} g(u') \geq g(u).$$

For sequences, a lsc function g satisfies

$$\liminf_{u^\nu \to u} g(u^\nu) \geq g(u).$$

When U is "sequential" (e.g., a Banach space), this property is equivalent to lower semicontinuity. We denote

$$\operatorname{argmin} g := \{u \in U \mid g(u) = \inf_{u' \in U} g(u')\}.$$

Theorem 2.11. *Let $\mathcal{J}$ be an arbitrary index set, g a lsc function and $(g^j)_{j \in \mathcal{J}}$ a collection of lsc functions on a topological vector space U. Then,*

1. *for a continuous $F : V \to U$, $g \circ F$ is lsc.*
2. *for finite $\mathcal{J}$ and strictly positive $(\lambda^j)_{j \in \mathcal{J}}$, the sum $\sum_{j \in \mathcal{J}} \lambda^j g^j$ is lsc,*
3. *the supremum $u \mapsto \sup_{j \in \mathcal{J}} g^j(u)$ is lsc.*

Proof. Exercise. □

Lemma 2.12. *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lsc such that $g(x) \geq -\rho|x| - m$ for $\rho, m \in \mathbb{R}_+$. The functions*

$$g^\nu(x) := \inf_{x' \in \mathbb{R}^d} \{g(x') + \nu\rho|x - x'|\} \quad \nu \in \mathbb{N}$$

are $(\nu\rho)$-Lipschitz with $g^\nu(x) \geq -\rho|x| - m$ and as ν increases, they increase pointwise to g. If g is convex, each g^ν is convex.

Proof. For any $x_1, x_2 \in \mathbb{R}^n$,

$$\begin{aligned} g^\nu(x_1) &\leq \inf_{x'} \{g(x') + \nu\rho|x' - x_2| + \nu\rho|x_2 - x_1|\} \\ &= g^\nu(x_2) + \nu\rho|x_2 - x_1|. \end{aligned}$$

By symmetry, g^ν is $\nu\rho$-Lipschitz continuous. For every ν and $\epsilon > 0$, there is a y^ν such that

$$\begin{aligned} g^\nu(x) &\geq h(y^\nu) + \nu\rho|y^\nu - x| - \epsilon \\ &\geq -\rho|y^\nu| - m + \nu\rho|y^\nu - x| - \epsilon \\ &\geq -\rho|x| - m + (\nu - 1)\rho|y^\nu - x| - \epsilon. \end{aligned}$$

Thus, either $g^\nu(x) \to \infty$ or $y^\nu \to x$ as $\nu \to \infty$. In the latter case, $\liminf g^\nu(x) \geq g(x)$ by lower semicontinuity of g. □

Let g be a lsc convex function. Given $\bar{u} \in \operatorname{dom} g$, the function

$$G(\lambda, u) := \begin{cases} \lambda(g(\bar{u} + u/\lambda) - g(\bar{u})) & \text{if } \lambda > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

is positively homogeneous and convex on $\mathbb{R} \times U$ by Exercise 2.4.2. The function $G(\cdot, u)$ is a decreasing on $\mathbb{R}_+$, i.e.,

$$\lambda \mapsto \frac{g(\bar{u} + \lambda u) - g(\bar{u})}{\lambda}$$

is increasing on $\mathbb{R}_+$. The function

$$u \mapsto g'(\bar{u}; u) = \lim_{\lambda \searrow 0} \frac{g(\bar{u} + \lambda u) - g(\bar{u})}{\lambda}$$

gives the *directional derivative* of g at $\bar{u}$. We have

$$g'(\bar{u}; u) = \inf_{\lambda \searrow 0} \frac{g(\bar{u} + \lambda u) - g(\bar{u})}{\lambda},$$

so $g'(\bar{u}, \cdot)$ is positively homogeneous and convex by Exercise 2.4.2. The function

$$g^\infty(u) = \lim_{\lambda \nearrow \infty} \frac{g(\bar{u} + \lambda u) - g(\bar{u})}{\lambda}$$

is called the *recession function* of g. Note that g^∞ is independent of the choice $\bar{u}$, and

$$g^\infty(u) = \sup_{\lambda > 0} \frac{g(\bar{u} + \lambda u) - g(\bar{u})}{\lambda},$$

so g^∞ is positive homogeneous and convex. Since g is lsc, g^∞ is lsc as well.

Theorem 2.13. *For a proper lsc convex function g, the function*

$$G(\lambda, u) := \begin{cases} \lambda(g(\bar{u} + u/\lambda) - g(\bar{u})) & \text{if } \lambda > 0, \\ g^\infty(u) & \text{if } \lambda = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

is lsc.

Proof. Exercise. □

For a convex C, the set

$$C^\infty := \{x \mid x' + \lambda x \in C \ \forall x' \in C, \ \lambda > 0\}$$

is called the *recession cone* of C. For a closed C, we have $\delta_C^\infty = \delta_{C^\infty}$, so the recession cone is closed for a closed C. This is a consequence of the following lemma.

Lemma 2.14. *For a closed convex C in a topological vector space,*

$$C^\infty := \{u \mid \bar{u} + \lambda u \in C \; \forall \lambda > 0\}$$

for any $\bar{u} \in C$.

Proof. It suffices show that the right side is a subset of C^∞. Let $u \neq 0$ and $\bar{u} \in C$ be such that $\bar{u} + \lambda u \in C$ for all $\lambda > 0$. Let $u' \in C$ and $\lambda' > 0$. For any $\lambda \geq \lambda'$,

$$u' + \lambda' u + \frac{\lambda'}{\lambda}(u - u') = (1 - \frac{\lambda'}{\lambda})u' + \frac{\lambda'}{\lambda}(\bar{u} + \lambda u) \in C$$

by convexity. Since C is closed, letting $\lambda \nearrow \infty$ gives $u' + \lambda' u \in C$. $\square$

Exercise 2.5.1. *If (x^ν) is a sequence in a closed convex C, $\lambda^\nu \searrow 0$ and $\lambda^\nu x^\nu \to \bar{x}$, then $\bar{x} \in C^\infty$.*

In a topological vector space, a set C is bounded if for any neighborhood A of the origin, $C \subset \lambda A$ for some $\lambda > 0$. In a normed space, like $\mathbb{R}^n$, this means that C is contained in some ball.

Theorem 2.15. *A convex set C in $\mathbb{R}^d$ is bounded if and only if $(\operatorname{cl} C)^\infty = \{0\}$.*

Proof. Exercise. $\square$

Remark 2.16. *In a general LCTVS, a closed set C need not be bounded even though $C^\infty = \{0\}$. Consider, e.g. $U = L^\infty$, the space of essentially bounded random variables equipped with a topology generated by the essential supremum norm. The set $C = \{u \in L^\infty \mid E|u| \leq 1\}$ is closed (this fact will follow from later facts proved in the course) and $C^\infty = \{0\}$ (this is easy to verify). If there are non-null $A^\nu \in \mathcal{F}$ with $P(A^\nu) \searrow 0$, then $u^\nu := 1_{A^\nu}/P(A^\nu)$ belongs to C but $\|u^\nu\|_{L^\infty} = P(A^\nu) \nearrow \infty$, so C is not bounded.*

Recall that in $\mathbb{R}^d$, a set is compact if and only if it is bounded and closed.

Theorem 2.17. *Let $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a proper lsc convex function. For any α with $\operatorname{lev}_{\leq\alpha} g \neq \emptyset$, we have*

$$(\operatorname{lev}_{\leq\alpha} g)^\infty = \operatorname{lev}_{\leq 0} g^\infty.$$

Moreover, $\operatorname{lev}_{\leq\alpha} g$ is bounded (and hence compact) for every α if and only if

$$\{x \mid g^\infty(x) \leq 0\} = \{0\}.$$

In this case, $\operatorname{argmin} g$ is nonempty and compact.

Proof. Exercise. $\square$

Exercise 2.5.2. *Assume that $g : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is a proper lsc convex function such that*

$$L := \{x \in \mathbb{R}^n \mid g^\infty(x, 0) \leq 0\}$$

is a linear space. Then $g(x + x', u) = g(x, u)$ for every $x' \in L$.

Theorem 2.18. *Assume that $g : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is a proper lsc convex function such that*

$$L := \{x \in \mathbb{R}^n \mid g^\infty(x, 0) \leq 0\}$$

is a linear space. Then the infimum in

$$p(u) := \inf_{x \in \mathbb{R}^n} g(x, u)$$

is attained, p is a proper lsc convex function and

$$p^\infty(u) = \inf_{x \in \mathbb{R}^n} g^\infty(x, u).$$

Proof. To prove that p that is lsc, it suffices to show that $\operatorname{lev}_{\leq\beta} p \cap \mathbb{B}$ is compact (and hence closed) for every closed ball $\mathbb{B}$. That $g(x+x') = g(x)$ for all $x' \in L$ is left as an exercise. Let $L^\perp := \{x \in \mathbb{R}^n \mid x \cdot x' = 0 \; \forall \; x' \in L\}$ and $\bar{g} = g + \delta_{L^\perp \times \mathbb{B}}$ so that $p(u) + \delta_{\mathbb{B}}(u) = \bar{p}(u) := \inf_{x \in \mathbb{R}^n} \bar{g}(x, u)$.

For a proper lsc convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we have that $\operatorname{lev}_{\leq\beta} f$ is bounded (and hence compact) for every β if and only if $\{x \mid f^\infty(x) \leq 0\} = \{0\}$ (an exercise). Applying this to $x \mapsto \bar{g}(x, u)$, this function has inf-compact level sets and thus the infimum in the definition of $\bar{p}$ and in that of p is attained for every u. In particular, we have $\operatorname{lev}_{\leq\beta} p \cap \mathbb{B} = \Pi(\operatorname{lev}_{\leq\beta} \bar{g})$ where Π is the projection from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^m$. We have

$$\{(x, u) \mid \bar{g}^\infty(x, u) \leq 0\} = \{(x, u) \mid u = 0, x \in L^\perp, g^\infty(x, u) \leq 0\} = \{(0, 0)\},$$

so $\bar{g}$ has compact-level sets as well. Thus $\operatorname{lev}_{\leq\beta} p \cap \mathbb{B}$ is a projection of a compact set and hence compact.

The proof of the recession formula is left as an exercise. □

2.6 Continuity of convex functions

Given a set $A \subset U$,

$$\operatorname{core} A = \{u \in U \mid \forall u' \in U \; \exists \lambda : u + \lambda' u \in A \; \forall \lambda \in (0, \lambda)\}$$

is known as the *core* (or algebraic interior) of A and its elements are called *internal points*, not to be confused with interior points. We always have $\operatorname{int} A \subset \operatorname{core} A$.

Theorem 2.19. *If a convex function is bounded from above on an open set, then it is continuous throughout the core of its domain.*

Proof. Let g be a function that is bounded from above on a neighborhood O of $\bar{u}$. We show first that this implies that g is continuous at $\bar{u}$. Replacing g by $g(u + \bar{u}) - g(\bar{u})$, we may assume that $\bar{u} = 0$ and that $g(0) = 0$. Hence there exists $M > 0$ such that $g(u) \leq M$ for all $u \in O$.

Let $\epsilon > 0$ be arbitrary and choose $\lambda \in (0,1)$ with $\lambda M < \epsilon$. We have $g(\lambda u) < \epsilon$ for each $u \in O$ by convexity. Moreover

$$0 = g((1/2((-\lambda u) + (1/2)\lambda u) \leq (1/2)g(-\lambda u) + (1/2)g(\lambda u),$$

which implies that $-g(-\lambda u) \leq g(\lambda u)$ for all $u \in O \cap (-O)$. Thus, $|g(u)| < \epsilon$ for all $u \in \alpha(O \cap (-O))$, so g is continuous at the origin.

Assume now that g is bounded from above on an open set A, i.e., there is M such that $g(u) \leq M$ for each $u \in A$. By above, it suffices to show that g is bounded from above at each $u \in \operatorname{core} \operatorname{dom} g$. Let $u' \in A$. There is $\bar{u} \in \operatorname{dom} g$ and $\lambda \in (0,1)$ with $u = \lambda\bar{u} + (1-\lambda)u'$. We have

$$g(\lambda\bar{u} + (1-\lambda)\tilde{u})) \leq \lambda g(\bar{u}) + (1-\lambda)M \quad \forall \tilde{u} \in A,$$

so g is bounded from above on a open neighborhood $\lambda\bar{u} + (1-\lambda)A$ of u. □

In $\mathbb{R}^d$, geometric intuition suggests that a convex function is continuous on the core of its domain. This idea extends to lsc convex functions on a barreled space. The LCTVS space U is *barreled* if every closed convex symmetric absorbing set is a neighborhood of the origin[2]. A set C is called *absorbent* if $\bigcup_{\alpha \in \mathbb{R}_+}(\alpha C) = U$. A set is absorbent if and only if the origin belongs to its core. For example, every Banach space is barreled[3].

Lemma 2.20. *Let $A \subset U$ be a convex set. Then* $\operatorname{int} A = \operatorname{core} A$ *under any of the following conditions:*

1. *U is finite dimensional,*
2. $\operatorname{int} A \neq \emptyset$*,*
3. *U is barreled and A is closed.*

Proof. We leave the first case as an exercise. To prove the second, it suffices to show that, for $u \in \operatorname{core} A$, we have $u \in \operatorname{int} A$. Let $u' \in \operatorname{int} A$. There is $\lambda \in (0,1)$ and $\bar{u} \in A$ with $u = \lambda\bar{u} + (1-\lambda)u'$. Now

$$u + (1-\lambda)(\operatorname{int} A - u') = \lambda\bar{u} + (1-\lambda)\operatorname{int} A \subset A$$

where the left side is an open neighborhood of u.

To prove the last claim, let U be barreled and A closed. Again, it suffices to show that, for $u \in \operatorname{core} A$, we have $u \in \operatorname{int} A$. Let $B = A - u$. Now $0 \in \operatorname{core} B$, so $0 \in \operatorname{core}(B \cap (-B))$. Thus $(B \cap (-B))$ is a closed convex symmetric absorbing set and hence it is a neighborhood of the origin. Thus $0 \in \operatorname{int} B$ and $x \in \operatorname{int} A$. □

[2] It is an exercise to show that in a LCTVS, every neighborhood of the origin is absorbing and contains a closed convex symmetric neighborhodhood of the origin.

[3] An application of the Baire category theorem: if $U = \bigcup_{n \in \mathbb{N}}(nC)$ for a closed C, then $\operatorname{int} C \neq \emptyset$

Theorem 2.21. *A convex function g is continuous on* $\operatorname{core}\operatorname{dom} g$ *in the following situations:*

1. *U is finite dimensional*
2. *U is barreled and g is lower semicontinuous.*

Proof. We leave the first part as an exercise. To prove the second, let $u \in \operatorname{core}\operatorname{dom} g$ and $\alpha > g(u)$. For $u' \in U$, the function $\lambda \mapsto g(u + \lambda u')$ is continuous at the origin by the first part, so $u \in \operatorname{core}\operatorname{lev}_\alpha g$. By Lemma 2.20 and lower semicontinuity of g, $\operatorname{int}\operatorname{lev}_\alpha g \neq \emptyset$. Thus continuity follows from Theorem 2.19. □

2.7 Convex conjugates

From now on, we assume that U and Y are vector spaces that are in separating duality under the bilinear form

$$\langle u, y \rangle.$$

That the bilinear form is separating means that for every $u \neq u'$, there is $y \in Y$ with $\langle u - u', y \rangle \neq 0$. On U the *weak* topology $\sigma(U, Y)$ is the weakest locally convex topology under which each

$$u \mapsto \langle u, y \rangle$$

is continuous. That is, $\sigma(U, Y)$ is generated by sets of the form

$$\{u \in U \mid |\langle u, y \rangle| < \alpha\}$$

where $\alpha > 0$ and $y \in Y$. Under $\sigma(U, Y)$, U is a locally convex topological space. The Mackey topology $\tau(U, Y)$ is the strongest locally convex topology under which each continuous linear functional can be identified with an element of Y. The Mackey topology is generated by sets of the form

$$\{u \in U \mid \sup_{y \in K} \langle u, y \rangle < 1\} \tag{2.1}$$

where K is convex symmetric and $\sigma(Y, U)$-compact.

Turning the idea around, when U is a locally convex topological vector space, a natural choice for Y is the *dual space* of continuous linear functionals on U. By Theorem 2.6, the bilinear form is separating. Especially for Banach spaces, $\sigma(U, Y)$ is called the weak topology and $\sigma(Y, U)$ the weak*-topology, and the Mackey topology $\tau(U, Y)$ coincides with the norm topology. We call both these topologies simply weak topologies, when the spaces in question are clear. When $U = \mathbb{R}^d$, we always choose $Y = \mathbb{R}^d$ and the bilinear form as the usual inner product.

Example 2.22. *Recall that, for $p \in [1, \infty)$, the Lebesque space $L^p := L^p(\Omega, \mathcal{F}, P)$ is a Banach space under the norm*

$$\|u\| := (E|u^p|)^{1/p}.$$

For $p > 1$, its continuous dual can be identified with $L^{p'}$ for p' satisfying $1/p + 1/p' = 1$. For $p = 1$, we set $p' = \infty$, and the continuous dual of L^1 is the space $L^\infty = L^\infty(\Omega, \mathcal{F}, P)$ of essentially bounded random variables. For all $p = [1, \infty)$, the bilinear form between L^p and $L^{p'}$ is given by

$$\langle u, y\rangle := E[u \cdot y].$$

Note, however, than when L^∞ is equipped with the essential supremum norm, its continuous dual is not L^1 (but the space of finitely additive measures on $(\Omega, \mathcal{F})$). In particular, $\tau(L^\infty, L^1)$ is a weaker topology than the one generated by the essential supremum norm.

Given an extended real-valued function g on U, its *conjugate* $g^* : Y \to \overline{\mathbb{R}}$ is

$$g^*(y) := \sup_{u \in \mathcal{U}}\{\langle u, y\rangle - g(u)\}.$$

The function g^* is also known as *Legendre-Fenchel transform*, *polar function*, or *convex conjugate* of g. Since g^* is a supremum of lower semicontinuous functions, g^* is a lower semicontinuous function on Y. The *Fenchel inequality*

$$g(u) + g^*(y) \geq \langle u, y\rangle$$

follows directly from the definition of the convex conjugate. In the exercises, we will familiarize ourselves with this transformation by calculating conjugates of convex functions defined on $\mathbb{R}^d$.

The *biconjugate* of g is the function

$$g^{**}(c) = \sup_y\{\langle u, y\rangle - g^*(y)\}.$$

By the Fenchel inequality, we always have

$$g \geq g^{**}.$$

The following *biconjugate theorem* is the fundamental theorem on convex conjugates. The *lower semicontinuous hull* $\operatorname{lsc} g$ is a function defined via

$$\operatorname{epi}(\operatorname{lsc} g) := \operatorname{cl} \operatorname{epi} g$$

while the *convex hull* $\operatorname{co} g$ is a function defined by

$$\operatorname{epi}(\operatorname{co} g) := \operatorname{co} \operatorname{epi} g$$

The *closure* of a convex function $g : U \to \overline{\mathbb{R}}$ is the function $\operatorname{cl} g : U \to \overline{\mathbb{R}}$ defined by

$$\operatorname{cl} g = \begin{cases} \operatorname{lsc} g & \text{if } \operatorname{lsc} g(u) > -\infty \text{ for all } u \in U, \\ -\infty & \text{otherwise.} \end{cases}$$

A function $g : U \to \overline{\mathbb{R}}$ is *closed* at $u \in U$ if $g(u) = (\operatorname{cl} g)(u)$. A function is *closed* if it is closed at every point.

Theorem 2.23 (Biconjugate theorem)**.** *Given a function $g : U \to \overline{\mathbb{R}}$,*

$$g^{**} = \operatorname{cl} \operatorname{co} g.$$

Proof. We have $(u, \alpha) \in \operatorname{epi} g^{**}$ if and only if

$$\alpha \geq \langle u, y \rangle - \beta \quad \forall (y, \beta) \in \operatorname{epi} g^*.$$

Here $(y, \beta) \in \operatorname{epi} g^*$ if and only if $g(u) \geq \langle u, y \rangle - \beta$ for every u. Thus, $\operatorname{epi} g^{**}$ is the intersection of the epigraphs of all continuous affine functionals dominated by g.

On the other hand, by Theorem 2.7, $\operatorname{epi} \operatorname{lsc} \operatorname{co} g$ is the intersection of all closed half-spaces

$$H_{y,\beta,\gamma} := \{(u, \alpha) \in U \times \mathbb{R} \mid \langle u, y \rangle + \alpha\beta \leq \gamma\}$$

containing $\operatorname{epi} g$. We have $(\operatorname{lsc} \operatorname{co} g)(u) > -\infty$ for every $u \in U$ if and only if one of the half-spaces has $\beta \neq 0$, or in other words, there is an affine function h_0 dominated by g.

It thus suffices to show that if there is a half-space $H_{y,\beta,\gamma}$ containing $\operatorname{epi} g$ but not a point $(\bar{u}, \bar{\alpha})$, then there is an affine function h such that $g \geq h$ but $h(\bar{u}) > \bar{\alpha}$. If $\operatorname{epi} g \subseteq H_{y,\beta,\gamma}$, then necessarily $\beta \leq 0$. If $\beta < 0$, then the function $h(u) = \langle u, y/(-\beta) \rangle + \gamma/\beta$ will do. If $\beta = 0$, then $\operatorname{dom} g$ is contained in $\{u \,|\, \langle u, y \rangle \leq \gamma\}$ while $\bar{u}$ is not. It follows that g dominates the affine function $h(u) = h_0(u) + \lambda(\langle u, y \rangle - \gamma)$ for any $\lambda \geq 0$. Since $\langle \bar{u}, y \rangle > \gamma$, we have $h(\bar{u}) > \bar{\alpha}$ for λ large enough. □

Exercise 2.7.1. *A proper convex function is $\sigma(U, Y)$-lsc if and only if it is $\tau(U, Y)$-lsc.*

Given a set $C \subset U$, the function

$$\sigma_C(y) = \sup_{u \in C} \langle u, y \rangle$$

is known as the *support function* of C,

$$j_C(u) := \inf_{\lambda > 0} \{\lambda \mid u \in \lambda C\}$$

as the *gauge* of C, and the set

$$C^\circ := \{y \mid \sigma_C(y) \leq 1\}$$

as the *polar* of C. Note that σ_C is the conjugate of the *indicator function*

$$\delta_C(u) = \begin{cases} 0 & \text{if } u \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

In the following theorem, $\operatorname{cl} C = C^{\circ\circ}$ is known as the *bipolar theorem.*

Theorem 2.24. *If a convex set C contains the origin, then*

$$\begin{aligned} \sigma_C &= j_{C^\circ}, \\ j_C^* &= \delta_{C^\circ}, \\ \operatorname{cl} C &= C^{\circ\circ}. \end{aligned}$$

Proof. Exercise. □

Theorem 2.25. *The set $C \subset U$ is a Mackey neighborhood of the origin if and only if $\{y \in Y \mid \sigma_C(y) \le \alpha\}$ is weakly compact for some $\alpha > 0$. In this case, $\{y \in Y \mid \sigma_C(y) \le \alpha\}$ is weakly compact for all α. In particular, C is a Mackey neighborhood of the origin if and only if C° is weakly compact.*

Proof. Let C be a Mackey neighborhood of the origin. By (2.1), $K^\circ \subset C$ for some convex weakly compact K for which

$$\{y \in Y \mid \sigma_C(y) \le \alpha\} \subset \{y \in Y \mid \sigma_{K^\circ}(y) \le \alpha\} = \alpha K^{\circ\circ} = \alpha K.$$

Thus the closed set on the left side belongs to a weakly compact set and is thus compact for all $\alpha > 0$.

To prove the converse, fix $\alpha > 0$ with $\alpha C^\circ = \{y \in Y \mid \sigma_C(y) \le \alpha\}$ weakly compact. The convex symmetric set $K := \operatorname{co}(\alpha C^\circ \cup (-\alpha C^\circ))$ is weakly compact as well (an exercise). By (2.1), K° is a Mackey neighborhood of the origin. Since $C^\circ \subset K/\alpha$, we have $\alpha K^\circ \subset C^{\circ\circ} = \operatorname{cl} C$, so C is a neighborhood of the origin. □

Theorem 2.26. *Let g be a proper convex lower semicontinuous function on U. The following are equivalent:*

1. *g is bounded from above on a $\tau(U,Y)$- neighborhood of $\bar{u}$,*
2. *for every $\alpha \in \mathbb{R}$, $\{y \in Y \mid g^*(y) - \langle \bar{u}, y\rangle \le \alpha\}$ is $\sigma(Y,U)$-compact.*

Here 1 implies 2 even when g is not lsc.

Proof. By translations, we may assume that $\bar{u} = 0$ and $g(0) = 0$. To prove that 1. implies 2., note that we have (conjugate inverts the order)

$$g(u) \le \gamma + \delta_O(u) \quad \forall\, u \quad \iff \quad g^*(y) \ge \sigma_O(y) - \gamma \quad \forall\, y \in Y,$$

so, for any $\alpha \in \mathbb{R}$,

$$\{y \in Y \mid g^*(y) \leq \alpha\} \subset \{y \in Y \mid \sigma_O(y) \leq \alpha + \gamma\},$$

where the set on the right side is weakly compact when O is a Mackey neighborhood of the origin.

That 2. implies 1., we may again do translations so that $g^*(0) = \inf_y g^*(y) = 0$; the details are left as an exercise. Let $\gamma > 0$ and denote

$$K := \{y \in Y \mid g^*(y) \leq \gamma\}.$$

If $y \notin K$, we have

$$\begin{aligned}
j_K(y) &:= \inf_{\lambda>0}\{\lambda \mid y \in \lambda K\} \\
&= \inf_{\lambda>1}\{\lambda \mid y \in \lambda K\} \\
&= \inf_{\lambda>1}\{\lambda \mid g^*(y/\lambda) \leq \gamma\} \\
&\leq \inf_{\lambda>1}\{\lambda \mid g^*(y)/\lambda \leq \gamma\} \\
&= g^*(y)/\gamma.
\end{aligned}$$

If $y \in K$, we have $j_K(y) \leq 1$, so putting these together we get that $g^*(y) \geq \gamma j_K(y) - \gamma$. Conjugating, we get

$$g(u) \leq \delta_{K^\circ}(u/\gamma) + \gamma.$$

Therefore, g is bounded above in a neighborhood of the origin, since K° is the polar of a weakly compact set. □

2.7.1 Exercises

Exercise 2.7.2. *Let f be a convex lower semicontinous function on the real line. Convince yourself that, given a "slope" v, $f^*(v)$ is the smallest constant α such that the affine function $x \to vx - \alpha$ is majorized by f. What does this mean geometrically?*

Exercise 2.7.3. *Calculate the conjugates of the following functions on the real line:*

1. $f(x) = |x|$
2. $f(x) = \delta_{\mathbb{B}}(x)$, *where* $\mathbb{B} = \{x \mid |x| \leq 1\}$.
3. $f(x) = \frac{1}{p}|x|^p$, *for* $p > 1$.
4. $V(x) = (e^{ax} - 1)/a$.

Exercise 2.7.4. *Let V be a nondecreasing convex function on the real line. Analyze V^* using the geometric idea from the first exercise.*

1. *Is V^* positive?*
2. *Is V^* zero somewhere?*
3. *Is V^* monotone?*
4. *Where is V^* finite?*
5. *Is V^* necessarily finite at the origin?*

Hint: The answers depends on your choice of V.

2.8 Subgradients of convex functions

Assume again that $g : U \to \overline{\mathbb{R}}$ is convex. Given u such that $g(u)$ is finite, a vector $y \in Y$ is a *subgradient* of g at u if

$$g(u') \geq g(u) + \langle u' - u, y \rangle \quad \forall u' \in U.$$

The *subdifferential* $\partial g(u)$ is the set of all subgradients of g at u. Note that we avoided defining subgradients at points where the function is not finite.

Exercise 2.8.1. *We have $y \in \partial g(u)$ if and only if*

$$g(u) + g^*(y) = \langle u, y \rangle.$$

We say that g is *subdifferentiable* at u if $\partial g(u) \neq \emptyset$.

Exercise 2.8.2. *Assume that g is a differentiable convex function on the real line. Then $\partial g(u) = \{g'(u)\}$, the derivative of g at u. Give an expression for g^∞ in terms of the derivative.*

Exercise 2.8.3. *Give an example of a proper lsc convex extended real-valued function on the real line that is not subdifferentiable at a point in its domain.*

Theorem 2.27. *Assume that g is proper and bounded from above in a neighborhood of u. Then $\partial g(u)$ is nonempty and weakly compact, and the directional derivate $g'(u, \cdot)$ is the support function of $\partial g(u)$.*

Proof. Exercise. Hint: show first that $g'(u, \cdot)$ is bounded above in a neighborhood of the origin. □

Remark 2.28. *Let U be a Frechet spaced (e.g., a Banach space) and $g : U \to \overline{\mathbb{R}}$. If g lsc proper convex and* $\operatorname{aff} \operatorname{dom} g$ *is closed, then g is either continuous and subdifferentiable or identically $-\infty$ throughout* $\operatorname{rcore} \operatorname{dom} g$. *Here* $\operatorname{rcore} \operatorname{dom} g$ *is the core relative to* $\operatorname{aff} \operatorname{dom} g$, *the smallest contain containing* $\operatorname{dom} g$. *We omit the proof.*

Theorem 2.29. *For a proper convex g, we have $(g^*)^\infty = \sigma_{\operatorname{dom} g}$. If g is also lsc, then $g^\infty = \sigma_{\operatorname{dom} g^*}$.*

Proof. Exercise. □

2.9 Conjugate duality in optimization

Consider a general convex optimization problem

$$\text{minimize} \quad f(x) \quad \text{over}\, x \in X, \tag{2.2}$$

where f is an extended real-valued convex function on a vector space X. We assume throughout that X is in separating duality with a vector space V. Given another pair (U, Y) of vector spaces in separating duality, we say, following [**?**] and [**?**], that a convex function $F : X \times U \to \overline{\mathbb{R}}$ is a *Rockafellian* for problem (2.2) if

$$f(x) = F(x, \bar{u}) - \langle x, \bar{v} \rangle$$

for some $\bar{u} \in U$ and $\bar{v} \in V$. Problem (2.2) can then be written as

$$\text{minimize} \quad F(x, \bar{u}) - \langle x, \bar{v} \rangle \quad \text{over}\, x \in X.$$

Within the duality framework described below, this will be called the *primal problem*. The *dual problem* associated with F is

$$\text{maximize} \quad \langle \bar{u}, y \rangle - F^*(\bar{v}, y) \quad \text{over}\, y \in Y.$$

The primal *optimum value function* is defined by

$$\varphi_v(u) := \inf_{x \in X} \{F(x, u) - \langle x, v \rangle\}$$

and the dual optimum value function by

$$\gamma_u(v) := \inf_{y \in Y} \{F^*(v, y) - \langle u, y \rangle\}.$$

By Theorem 2.8, these functions are convex. By Fenchel's inequality,

$$F(x, u) + F^*(v, y) \geq \langle x, v \rangle + \langle u, y \rangle \quad \forall (x, u) \in X \times U,\ (v, y) \in V \times Y,$$

so

$$\varphi_v(u) \geq -\gamma_u(v) \quad \forall u \in U, v \in V.$$

If $\varphi_{\bar{v}}(\bar{u}) > -\gamma_{\bar{u}}(\bar{v})$, a *duality gap* is said to exist. Note that $\varphi_v^*(y) = F^*(v, y)$ and $\gamma_u^*(x) = F^{**}(x, u)$, where $F^{**} = \operatorname{cl} F$, by Theorem 2.23. In particular, the dual problem can be written as

$$\text{maximize} \quad \langle \bar{u}, y \rangle - \varphi_{\bar{v}}^*(y) \quad \text{over}\, y \in Y.$$

The properties of conjugates and subgradients from the previous section thus imply that the absence of a duality gap and existence of dual solutions come down to lower semicontinuity and subdifferentiability of $\varphi_{\bar{v}}$ at $\bar{u}$; see Theorem 2.31 and Theorem 2.32 below.

The *Lagrangian* associated with F is the convex-concave function on $X \times Y$ given by

$$L(x,y) := \inf_{u \in U}\{F(x,u) - \langle u, y\rangle\}.$$

Clearly, the conjugate of F can be expressed as

$$F^*(v,y) = \sup_{x \in X}\{\langle x, v\rangle - L(x,y)\}.$$

The *Lagrangian minimax problem* to find a saddle value and/or a saddle point of the convex-concave function

$$L_{\bar{v},\bar{u}}(x,y) := L(x,y) - \langle x, \bar{v}\rangle + \langle \bar{u}, y\rangle.$$

We always have

$$\inf_x \sup_y L_{\bar{v},\bar{u}}(x,y) \geq \sup_y \inf_x L_{\bar{v},\bar{u}}(x,y).$$

When the equality holds, the common value is called the *minimax* or the *saddle value* of $L_{\bar{v},\bar{u}}$. A pair $(\bar{x},\bar{y})$ is called a *saddle point* of $L_{\bar{v},\bar{u}}$ if

$$L_{\bar{v},\bar{u}}(x,\bar{y}) \geq L_{\bar{v},\bar{u}}(\bar{x},\bar{y}) \geq L_{\bar{v},\bar{u}}(\bar{x},y) \quad \forall x \in X, y \in Y.$$

When a saddle point $(\bar{x},\bar{y})$ exists, the saddle value exists and equals $L_{\bar{v},\bar{u}}(\bar{x},\bar{y})$.

We have

$$\langle \bar{u}, y\rangle - F^*(\bar{v},y) = \inf_x L_{\bar{v},\bar{u}}(x,y) \tag{2.3}$$

so the dual problem can be viewed as the maximization-half of the Lagrangian minimax problem. When F is closed in u in the sense $F(x,\cdot) = \operatorname{cl} F(x,\cdot)$ for all x, the biconjugate theorem gives

$$F(x,\bar{u}) - \langle x, \bar{v}\rangle = \sup_y L_{\bar{v},\bar{u}}(x,y),$$

so the primal problem is the minimization-half of the minimax problem. In general,

$$(\operatorname{cl}_u F)(x,\bar{u}) - \langle x, \bar{v}\rangle = \sup_y L_{\bar{v},\bar{u}}(x,y), \tag{2.4}$$

where $(\operatorname{cl}_u F)(x,\cdot) := \operatorname{cl} F(x,\cdot)$. Clearly,

$$\operatorname{cl} F \leq \operatorname{cl}_u F \leq F.$$

Recalling that $\varphi_v^*(y) = F^*(v,y)$ and $\gamma_u^*(x) = (\operatorname{cl} F)(x,u)$, Theorem 2.23 yields the following.

Lemma 2.30. *We have*

$$\begin{aligned}
(\operatorname{cl}\varphi_v)(u) &= \sup_y\{\langle u,y\rangle - \varphi_v^*(y)\}\\
&= \sup_y\{\langle u,y\rangle - F^*(v,y)\}\\
&= -\gamma_u(v)\\
&\le -(\operatorname{cl}\gamma_u)(v)\\
&= \inf_x\{\gamma_u^*(x) - \langle x,v\rangle\}\\
&= \inf_x\{(\operatorname{cl}F)(x,u) - \langle x,v\rangle\}\\
&\le \inf_x\{(\operatorname{cl}_u F)(x,\bar u) - \langle x,v\rangle\}\\
&\le \inf_x\{F(x,u) - \langle x,v\rangle\}\\
&= \varphi_v(u).
\end{aligned}$$

The following is an immediate consequence of Lemma 2.30 and equations (2.3) and (2.4).

Theorem 2.31. *The implications* $1 \Leftrightarrow 2 \Rightarrow 3 \Rightarrow 4$ *hold among the following conditions:*

1. *there is no duality gap;*
2. $\varphi_{\bar v}$ *is closed at* $\bar u$*;*
3. $L_{\bar v,\bar u}$ *has a saddle value;*
4. $\gamma_{\bar u}$ *is closed at* $\bar v$*.*

If F *is closed in* u*, then* $1 \Leftrightarrow 2 \Leftrightarrow 3$*. If* F *is closed, then* $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$*.*

Theorem 2.32. *Assume that* $\varphi_{\bar v}(\bar u) < \infty$*. The implications* $1 \Leftrightarrow 2 \Rightarrow 3 \Rightarrow 4$ *hold among the following conditions:*

1. *there is no duality gap and* y *solves the dual;*
2. *either* $y \in \partial\varphi_{\bar v}(\bar u)$ *or* $\varphi_{\bar v}(\bar u) = -\infty$*;*
3. $\inf_x \sup_y L_{\bar v,\bar u}(x,y) = \inf_x L_{\bar v,\bar u}(x,y)$*;*
4. $\gamma_{\bar u}$ *is closed at* $\bar v$ *and* y *solves the dual.*

If F *is closed in* u*, then* $1 \Leftrightarrow 2 \Leftrightarrow 3$*. If* F *is closed, then* $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$*.*

Proof. Condition 1 means that either $\varphi_{\bar v}(\bar u) + \varphi_{\bar v}^*(y) = \langle \bar u, y\rangle$ or $\varphi_{\bar v}(\bar u) = -\infty$. Indeed, in the latter case, every $y \in Y$ solves the dual. This proves the equivalence of 1 and 2. The remaining claims follow from Lemma 2.30 and equations (2.3) and (2.4). □

Theorem 2.33. *The implications* $1 \Leftrightarrow 2 \Leftrightarrow 3 \Rightarrow 4 \Leftrightarrow 5 \Rightarrow 6 \Leftrightarrow 7$ *hold among the following conditions:*

1. *There is no duality gap, x solves the primal, y solves the dual and both problems are feasible;*
2. $y \in \partial\varphi_{\bar{v}}(\bar{u})$ *and x solves the primal;*
3. $(\bar{v}, y) \in \partial F(x, \bar{u})$*;*
4. (x, y) *is a saddle point of* $L_{\bar{u},\bar{v}}$*;*
5. $\bar{v} \in \partial_x L(x, y)$ *and* $\bar{u} \in \partial_y[-L](x, y)$*;*
6. $x \in \partial\gamma_{\bar{u}}(\bar{v})$ *and y solves the dual;*
7. $(x, \bar{u}) \in \partial F^*(\bar{v}, y)$*.*

If F is closed in u, then $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4 \Leftrightarrow 5$*. If F is closed, then* $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4 \Leftrightarrow 5 \Leftrightarrow 6 \Leftrightarrow 7$*.*

Proof. By Lemma 2.30, 1 means that

$$\langle \bar{u}, y\rangle - \varphi^*_{\bar{v}}(y) = \langle \bar{u}, y\rangle - F^*(\bar{v}, y) = F(x, \bar{u}) - \langle x, \bar{v}\rangle = \varphi_{\bar{v}}(\bar{u}),$$

which is equivalent to both 2 and 3. Condition 5 is simply a reformulation of 4. By Lemma 2.30, 7 means that

$$\langle \bar{u}, y\rangle - F^*(\bar{v}, y) = -\gamma_{\bar{u}}(\bar{v}) = \gamma^*_{\bar{u}}(x) - \langle x, \bar{v}\rangle = (\operatorname{cl} F)(x, \bar{u}) - \langle x, \bar{v}\rangle.$$

Since $\gamma^*_{\bar{u}}(x) = (\operatorname{cl} F)(x, \bar{u})$, this is equivalent to 6. By Lemma 2.30, 1 implies

$$\langle \bar{u}, y\rangle - F^*(\bar{v}, y) = (\operatorname{cl}_u F)(x, \bar{u}) - \langle x, \bar{v}\rangle \tag{2.5}$$

which is 4. If $(\operatorname{cl}_u F) = F$, this is, by Lemma 2.30, equivalent to 1. The equality (2.5) clearly implies

$$\langle \bar{u}, y\rangle - F^*(\bar{v}, y) = (\operatorname{cl} F)(x, \bar{u}) - \langle x, \bar{v}\rangle$$

which is 7. If $\operatorname{cl} F = F$, this is, by Lemma 2.30, equivalent to 1. □

Condition 5 of Theorem 2.33 is sometimes referred to as "Kuhn-Tucker conditions", although condition 5 above is far more general than the original Kuhn-Tucker conditions which only apply to optimization problems with explicit equality and inequality constraints; see Example 2.36 below. It has now become customary to refer to the optimality conditions in Example 2.36 as *Karush-Kuhn-Tucker (KKT) conditions* to also credit the work of Karush who introduced the conditions a few years before Kuhn and Tucker. The generalization in condition 5 of Theorem 2.33 is due to Rockafellar. Accordingly, we will refer to condition 5 in Theorem 2.33 as the KKTR-conditions.

Combining Theorem 2.32 with general properties of convex functions on locally convex spaces, we obtain the following.

Lemma 2.34. *Assume that F is closed and proper and that, for every $\bar{v} \in V$, $\varphi_{\bar{v}}$ is either relatively Mackey continuous at $\bar{u} \in U$ or $\varphi_{\bar{v}}(\bar{u}) = -\infty$. Then, for every $\bar{v} \in V$, there exists y such that all the conditions in Theorem 2.32 hold and*

1. *$\gamma_{\bar{u}}$ is closed and proper and the infimum in its definition is attained,*
2. *the recession function of $\gamma_{\bar{u}}$ is given by*

$$\gamma_{\bar{u}}^{\infty}(v) = \inf_{y \in Y} \{(F^*)^{\infty}(v, y) - \langle \bar{u}, y \rangle\},$$

   where the infimum is attained.

In particular, $\gamma_{\bar{u}} = F(\cdot, \bar{u})^*$.

Proof. Omitted. □

Note that $\operatorname{dom} \varphi_v$ does not depend on v. We denote this set by $\operatorname{dom} \varphi$.

Theorem 2.35. *Let X and U be Fréchet and let V and Y, respectively, be the corresponding topological duals. Assume that F is closed and proper, $\bar{u} \in \operatorname{rcore} \operatorname{dom} \varphi$ and that $\operatorname{aff} \operatorname{dom} \varphi$ is closed. Then, for every $\bar{v} \in V$, $\varphi_{\bar{v}}$ is either relatively Mackey continuous at $\bar{u}$ or $\varphi_{\bar{v}}(\bar{u}) = -\infty$, there exists y such that all the conditions in Theorem 2.32 hold and*

1. *$\gamma_{\bar{u}}$ is closed and proper and the infimum in its definition is attained,*
2. *the recession function of $\gamma_{\bar{u}}$ is given by*

$$\gamma_{\bar{u}}^{\infty}(v) = \inf_{y \in Y} \{(F^*)^{\infty}(v, y) - \langle \bar{u}, y \rangle\},$$

   where the infimum is attained.

In particular, $\gamma_{\bar{u}} = F(\cdot, \bar{u})^*$.

Proof. Omitted. □

The condition $\bar{u} \in \operatorname{rcore} \operatorname{dom} \varphi$ in Theorem 2.35 holds if the primal problem is *strictly feasible* in the sense that there is an $\bar{x} \in X$ with $(\bar{x}, \bar{u}) \in \operatorname{rcore} \operatorname{dom} F$. If U is finite dimensional, then the closedness assumption in Theorem 2.35 is redundant. The closedness assumption holds also if $\bar{u} \in \operatorname{int} \operatorname{dom} \varphi$ which also implies $\bar{u} \in \operatorname{core} \operatorname{dom} \varphi$. The interiority condition generalizes classical Slater-type conditions which ask that $\bar{u} \in \operatorname{int} \operatorname{dom} F(\bar{x}, \cdot)$ for some $\bar{x} \in X$. In general, $\operatorname{aff} \operatorname{dom} \varphi$ is the projection of $\operatorname{aff} \operatorname{dom} F$ to U. The closedness of projections of linear spaces is a classical question in functional analysis and various sufficient conditions have been given.

Example 2.36 (Lagrangian duality)**.** *Let X be a Fréchet space, $U = \mathbb{R}^m$, f_j, $j = 0,\dots,l$ lsc convex and f_j, $j = l+1,\dots,m$ continuous affine functions on X. Consider the problem*

$$\begin{array}{lll} \text{minimize} & f_0(x) & \text{over}\, x \in X, \\ \text{subject to} & f_j(x) \le 0 & j = 1,\dots,l, \\ & f_j(x) = 0 & j = l+1,\dots,m. \end{array}$$

This fits the general duality framework with V the topological dual of X, $Y = \mathbb{R}^m$, the Rockafellian

$$F(x,u) = \begin{cases} f_0(x) & \text{if } f_j(x) + u_j \le 0 \text{ for } j = 1,\dots,l, \\ & \quad f_j(x) + u_j = 0 \text{ for } j = l+1,\dots,m, \\ +\infty & \text{otherwise,} \end{cases}$$

$\bar{v} = 0$ and $\bar{u} = 0$. The Lagrangian becomes

$$L(x,y) = \begin{cases} +\infty & \text{if } x \notin \cap_{j=0}^l \operatorname{dom} f_j, \\ f_0(x) + \sum_{j=1}^m y_j f_j(x) & \text{if } x \in \cap_{j=0}^l \operatorname{dom} f_j \text{ and } y \in \mathbb{R}^l_+, \\ -\infty & \text{otherwise.} \end{cases}$$

Assume that there exists a feasible $\bar{x} \in X$ such that $f_j(\bar{x}) < 0$ for $j = 1,\dots,l$ and either $m = l$ (no equality constraints) or $\bar{x} \in \cap_{j=0}^l \operatorname{rcore} \operatorname{dom} f_j$. Then the primal optimum value equals $\sup_y \inf_x L(x,y)$ where the supremum is attained. Moreover, an $x \in X$ solves the primal problem if and only if it is feasible and there exists $y \in \mathbb{R}^m$ such that

$$\partial_x [f_0 + \sum_{j=1}^m y_j f_j](x) \ni 0,$$
$$y_j \ge 0, \quad y_j f_j(x) = 0 \quad j = 1,\dots,l.$$

Proof. We prove only the case where there are no equality constraints. To prove the first claim, it suffices, by Theorem 2.35, to show that $0 \in \operatorname{rcore} \operatorname{dom} \varphi$, since $\operatorname{aff} \operatorname{dom} \varphi$ is automatically closed in $\mathbb{R}^m$. If there are no equality constraints, $\varphi(u)$ is bounded from above by the constant $f_0(\bar{x})$ on the set $\{u \in \mathbb{R}^m \mid f_j(\bar{x}) + u_j \le 0,\ j = 1,\dots,m\}$ which is a neighborhood of the origin in $\mathbb{R}^m$, so $0 \in \operatorname{rcore} \operatorname{dom} \varphi$. By Theorem 2.33, the first claim implies the second one. □

We close this section by some general remarks on saddle functions. The Lagrangian L is not necessarily closed in x but $-L$ is always closed in y. By the biconjugate theorem,

$$(\operatorname{cl}_x L)(x,y) = \sup_v \{\langle x,v\rangle - F^*(v,y)\},$$

where $(\mathrm{cl}_x L)$ is defined by $(\mathrm{cl}_x L)(\cdot, y) := \mathrm{cl}\, L(\cdot, y)$ for every $y \in Y$. When F is closed in u,

$$F(x,u) = \sup_y \{L(x,y) + \langle u, y\rangle\}$$

and, when F is closed,

$$F(x,u) = \sup_y \{(\mathrm{cl}_x L)(x,y) + \langle u, y\rangle\}.$$

The following can be found in [**?**].

Theorem 2.37. *If F is closed, then all saddle functions between $\mathrm{cl}_x L$ and L have the same saddle value and saddle points.*

Proof. Given a function $(x,y) \mapsto \tilde{L}(x,y)$, we define

$$(\mathrm{cl}_y \tilde{L})(x,\cdot) := -\,\mathrm{cl}(-\tilde{L})(x,\cdot) \quad \forall x \in X.$$

Since the infimums of a function and of its closure coincide, Theorem 2.23 gives

$$\begin{aligned}\sup_y \{\langle u, y\rangle + (\mathrm{cl}_y\, \mathrm{cl}_x L)(x,y)\} &= \sup_y \{\langle u, y\rangle + (\mathrm{cl}_x L)(x,y)\} \\ &= \sup_{v,y} \{\langle x, v\rangle + \langle u, y\rangle - F^*(v,y)\} \\ &= F(x,u).\end{aligned}$$

As already noted, $F(x,u) = \sup_y \{L(x,y) + \langle u, y\rangle\}$. Thus, by Theorem 2.23 again, $\mathrm{cl}_y\, \mathrm{cl}_x L = L$. It follows that

$$\inf_x L(x,y) = \inf_x (\mathrm{cl}_x L)(x,y) \quad \forall y \in Y$$

and

$$\sup_y L(x,y) = \sup_y (\mathrm{cl}_y\, \mathrm{cl}_x L)(x,y) = \sup_y (\mathrm{cl}_x L)(x,y) \quad \forall x \in X.$$

The claims now follow from the fact that saddle values and saddle points are characterized by the expressions on the left. □

Exercise 2.9.1. *Let $X = U = \mathbb{R}$. Find a lsc convex function $F : X \times U$ such that the primal and the dual problem have solutions but there is a duality gap.*

3 Normal integrands and integral functionals

Throughout, a finite dimensional space $\mathbb{R}^n$ is equipped with the usual topology and the usual Borel-σ-algebra $\mathcal{B}(\mathbb{R}^n)$. Given a measurable space $(\Omega, \mathcal{F})$, $\mathcal{L}^0(\mathbb{R}^n)$ is the space of measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

3.1 Random sets

Throughout $S : \Omega \rightrightarrows \mathbb{R}^d$ is a set-valued mapping, i.e., for every ω, $S(\omega) \subset \mathbb{R}^d$. The mapping S is *measurable* if the *preimage*

$$S^{-1}(O) := \{\omega \in \Omega \mid S(\omega) \cap O \neq \emptyset\}$$

of every open $O \subset \mathbb{R}^d$ is measurable, i.e. $S^{-1}(O) \in \mathcal{F}$ for every open O. The mapping S is (resp. *convex*, *cone*, etc.) *closed-valued* when $S(\omega)$ is (resp. *convex*, *conical*, etc.) closed for each $\omega \in \Omega$. The set

$$\operatorname{dom} S := \{\omega \mid S(\omega) \neq \emptyset\}$$

is the *domain* of S. Being the preimage of the whole space, it is measurable as soon as S is measurable. If S is measurable, then its *image-closure mapping* $\omega \mapsto \operatorname{cl} S(\omega)$ is measurable, since its preimages of open sets coincide with those of S. The function

$$d(x, A) = \inf_{x' \in A} |x - x'|$$

is the *distance mapping*. We denote the closed euclidean ball centered at x with radius r by $\mathbb{B}_r(x)$. When the ball is centered at the origin, we denote $\mathbb{B}_r$.

Theorem 3.1. *Let $S : \Omega \rightrightarrows \mathbb{R}^n$ be closed-valued. The following are equivalent.*

1. *S is measurable,*
2. *$S^{-1}(C) \in \mathcal{F}$ for every compact set C,*
3. *$S^{-1}(C) \in \mathcal{F}$ for every closed set C,*
4. *$S^{-1}(B) \in \mathcal{F}$ for every closed ball B,*
5. *$S^{-1}(O) \in \mathcal{F}$ for every open ball O,*
6. *$\{\omega \in \Omega \mid S(\omega) \subset O\} \in \mathcal{F}$ for every open O,*
7. *$\{\omega \in \Omega \mid S(\omega) \subset C\} \in \mathcal{F}$ for every closed C,*
8. *$\omega \mapsto d(x, S(\omega))$ is measurable for every $x \in \mathbb{R}^n$.*

Proof. Exercise. Assuming 1, for a compact C we have

$$S^{-1}(C) = \bigcap_{\nu} S^{-1}(C + \operatorname{int} \mathbb{B}_{1/\nu})$$

which proves 2. For a closed C, we have $C = \bigcup_{\nu}(C \cap \mathbb{B}_{\nu})$, so 3 implies 4. Clearly, 3 implies 4. For any open ball O,

$$O = \bigcup_{q,r} \{\mathbb{B}_r(r) \mid, \mathbb{B}_r(q) \subset O, (q,r) \in \mathbb{Q}^n, \mathbb{Q}_+\},$$

so 4 implies 5. Any open O is a countable union of open balls, so 5 implies 1. Since $\{\omega \mid S(\omega) \subseteq A\} = \Omega \backslash S^{-1}(A^C)$, 3 and 6 are equivalent, and 4 and 7 are equivalent. For any $r \in \mathbb{R}_+$, $d(x, S(\omega)) \leq r$ if and only if $S(\omega) \cap \mathbb{B}_r(x) \neq \emptyset$, so 4 and 8 are equivalent. □

For a set-valued mapping $S : \Omega \rightrightarrows \mathbb{R}^n$,

$$\operatorname{gph} S := \{(x, \omega) \in \mathbb{R}^n \times \Omega \mid x \in S(\omega)\}$$

is the *graph of* S.

Corollary 3.2. *The graph of a measurable closed-valued mapping is* $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-*measurable.*

Proof. Since S is closed, $x \in S(\omega)$ if and only if, for every $r \in \mathbb{Q}_+$, there is $q \in \mathbb{Q}^n$ such that $\omega \in S^{-1}(B_r(q))$, where $B_r(q)$ is an open ball centered at q with radius r containing x. Thus

$$\operatorname{gph} S = \bigcap_{r \in \mathbb{Q}_+} \bigcup_{q \in \mathbb{Q}^n} [S^{-1}(B_r(q)) \times B_r(q)],$$

where the right side is measurable. □

Measurability is preserved under many algebraic operations.

Theorem 3.3. *Let $\mathcal{J}$ be a countable index set and let each S^j, $j \in \mathcal{J}$, be a measurable set-valued mapping. Then*

1. $\omega \mapsto \bigcap_{j \in \mathcal{J}} S^j(\omega)$ *is measurable if each* S^j *is closed,*
2. $\omega \mapsto \bigcup_{j \in \mathcal{J}} S^j(\omega)$ *is measurable,*
3. $\omega \mapsto \sum_{j \in \mathcal{J}} \lambda^j S^j(\omega)$ *is measurable for finite* $\mathcal{J}$, *where* $\lambda^j \in \mathbb{R}$,
4. $\omega \mapsto (S^1(\omega), \ldots, S^j(\omega))$ *is measurable for finite* $\mathcal{J}$*; here we may allow* $S^j : \Omega \rightrightarrows \mathbb{R}^{d_j}$.

Proof. 4. Let $R(\omega) = (S^1(\omega), \dots, S^j(\omega))$. Every open set O in the product space is expressible as a union of rectangular open sets $\times_j O^j_\nu$. Thus $R^{-1}(O) = \bigcup_\nu (\cap_j (S^j)^{-1}(O^j_\nu))$, where each $(S^j)^{-1}(O^j_\nu)$ is measurable by the assumption.

3. Let $R(\omega) = \sum_{j=1}^J \lambda^j S^j(\omega)$. For an open O, the set

$$O' = \{(x^1, \dots, x^J) | \sum_{j=1}^{J} \lambda^\nu x^j \in O\}$$

is open in $\times_{j=1}^J \mathbb{R}^d$. Now

$$\begin{aligned} R^{-1}(O) &= \{\omega \mid (\sum_{j=1}^{J} \lambda^j S^j)(\omega) \cap O \neq \emptyset\} \\ &= \{\omega \mid (S^1(\omega), \dots, S^J(\omega)) \cap O' \neq \emptyset\}, \end{aligned}$$

where the set on the right-hand side is measurable by part 4.

2. $(\bigcup_{j\in\mathcal{J}} S^j)^{-1}(O) = \bigcup_{j\in\mathcal{J}} (S^j)^{-1}(O)$ for any open O.

1. Assume first that $\mathcal{J} = \{1, 2\}$. Take any compact $C \subset \mathbb{R}^d$, and denote $R^\nu(\omega) = S^\nu(\omega) \cap C$, then

$$\begin{aligned} (S^1 \cap S^2)^{-1}(C) &= \{\omega \mid S^1(\omega) \cap S^2(\omega) \cap C \neq \emptyset\} \\ &= \{\omega \mid 0 \in R^1(\omega) - R^2(\omega)\} \\ &= (R^1 - R^2)^{-1}(\{0\}). \end{aligned}$$

Here $R^1 - R^2$ is measurable by part 3; let us show that it is closed-valued as well. Since S^ν are closed-valued, R^ν are compact-valued, so $R^1 - R^2$ is compact valued. (an exercise). Hence $S^1 \cap S^2$ is measurable. The case of finite $\mathcal{J}$ follows from by induction.

Suppose finally that $\mathcal{J}$ is countable, $\mathcal{J} = \{1, 2, 3, \dots\}$. Denote $\tilde{S}^\mu = \bigcap_{\nu=1}^{\mu} S^\nu$. Note that $\bigcap_{\nu=1}^\infty S^\nu(\omega) = \bigcap_{\mu=1}^\infty \tilde{S}^\mu(\omega)$, and that $\tilde{S}^\mu$ are measurable by preceding. The proof is complete as soon as we show

$$(\bigcap S^\nu)^{-1}(C) = \bigcap_{\mu=1}^{\infty} (\tilde{S}^\mu)^{-1}(C).$$

If $\omega \in (\bigcap S^\nu)^{-1}(C)$, it is straight-forward to check that $\omega \in \bigcap_{\mu=1}^\infty (\tilde{S}^\mu)^{-1}(C)$. For the converse, take $\omega \in \bigcap_{\mu=1}^\infty (\tilde{S}^\mu)^{-1}(C)$. Since $(\tilde{S}^\mu(\omega) \cap C)_{\mu=1}^\infty$ is a nested sequence of nonempty compact sets, $\bigcap_{\mu=1}^\infty (\tilde{S}^\mu(\omega) \cap C) \neq \emptyset$. By $\bigcap_{\nu=1}^\infty S^\nu(\omega) = \bigcap_{\mu=1}^\infty \tilde{S}^\mu(\omega)$ this means that $\omega \in (\bigcap S^\nu)^{-1}(C)$. □

Given a set A, the *convex hull* $\operatorname{co} A$ of A is the smallest convex set containing A. Equivalently, $\operatorname{co} A$ is the set of all convex combinations of the points of A.

Corollary 3.4. *Assume that S is a measurable set-valued mapping. Then*

1. $\omega \mapsto \operatorname{co} S(\omega)$ *is measurable.*

Proof. The mapping $\operatorname{cl}\operatorname{co} S$ is the closure of a countable union of mappings of the form $\sum_{j\in\mathcal{J}} \lambda^j S$, where $\mathcal{J}$ is finite and $\lambda^i \geq 0$ are rational with $\sum_{j\in\mathcal{J}} \lambda^j = 1$. Thus $\operatorname{cl}\operatorname{co} S$ is measurable by Theorem 3.3, and so is $\operatorname{co} S$. □

Theorem 3.5. *If $M(\cdot,\omega) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is such that*

$$\operatorname{gph} M(\omega) := \{(x,u) \mid u \in M(x,\omega)\}$$

defines a measurable closed-valued mapping, then the following mappings are measurable,

1. $R(\omega) := M(S(\omega),\omega)$, *where $S : \Omega \rightrightarrows \mathbb{R}^n$ is measurable and closed-valued,*
2. $\omega \mapsto \{x \in \mathbb{R}^n \mid M(x,\omega) \cap S(\omega) \neq \emptyset\}$, *where $S : \Omega \rightrightarrows \mathbb{R}^m$ is measurable and closed-valued.*

Proof. Let $\Pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projection mapping. We have $R = \Pi \circ Q$ for $Q(\omega) := [S(\omega) \times \mathbb{R}^m] \cap \operatorname{gph} M(\omega)$, which is measurable by Theorem 3.1. Since $R^{-1}(O) = Q^{-1}(\Pi^{-1}(O))$, where $\Pi^{-1}(O)$ is open for any open O, R is measurable.

To prove 2, we have $\{x \mid M(x,\omega) \cap S(\omega) \neq \emptyset\} = \Gamma(S(\omega),\omega)$ for $\Gamma(\cdot,\omega) := M^{-1}(\cdot,\omega)$. Here $\operatorname{gph}\Gamma(\omega) = \{(x,u) \mid u \in M(x,\omega)\}$, so the result follows from 1. □

3.2 Normal integrands

A function $h : \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a *normal integrand* on $\mathbb{R}^n$ if

$$\operatorname{epi} h(\cdot,\omega) := \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \mid h(x,\omega) \leq \alpha\}$$

defines a measurable closed-valued mapping. A normal integrand is *convex* (positively homogeneous etc.), if, for all ω, $h(\cdot,\omega)$ is convex (positively homogeneous etc). The indicator function of a set-valued mapping S,

$$\delta_S(x,\omega) := \delta_{S(\omega)}(x) := \begin{cases} 0 & \text{if } x \in S(\omega), \\ +\infty & \text{otherwise} \end{cases}$$

defines a normal integrand on $\mathbb{R}^n$ if and only if S is closed-valued and measurable.

A function h is a *Caratheodory integrand* if $h(\cdot,\omega)$ is continuous for each $\omega \in \Omega$ and $h(x,\cdot)$ is measurable for each $x \in \mathbb{R}^n$.

Theorem 3.6. *A Caratheodory integrand is a normal integrand.*

Proof. Let $\{x^\nu \mid \nu \in \mathbb{N}\}$ be a dense set in $\mathbb{R}^d$ and define $\alpha^{\nu,q}(\omega) = h(x^\nu, \omega) + q$, where $q \in \mathbb{Q}_+$. Since $h(\cdot, \omega)$ is continuous, the set $\hat{O} = \{(x,\alpha) \mid h(x,\omega) < \alpha\}$ is open. For any $(x,\alpha) \in \operatorname{epi} h(\cdot,\omega)$ and for any open neighborhood O of (x,α), $O \cap \hat{O}$ is open and nonempty, and there exists $(x^\nu, \alpha^{\nu,q}) \in O \cap \hat{O}$, i.e., $\{(x^\nu, \alpha^{\nu,q}(\xi) \mid \nu \in \mathbb{N}, q \in \mathbb{Q}\}$ is dense in $\operatorname{epi} h(\cdot,\omega)$. Thus for any open set $O \in U \times \mathbb{R}$,

$$\{\omega \mid \operatorname{epi} h(\omega) \cap O \neq \emptyset\} = \bigcup_{\nu,q} \{\omega \mid (x^\nu, \alpha^{\nu,q}(\omega)) \in O\}$$

is measurable. □

Let $S_h(\omega) := \operatorname{epi} h(\cdot,\omega)$ and

$$S_h^o(\omega) := \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \mid h(x,\omega) < \alpha\}.$$

Since preimages of open sets under S_h and S_h^o are the same, one is measurable if and only if the other is so.

Theorem 3.7. *For a normal integrand h on $\mathbb{R}^n$ and $\beta \in \mathcal{L}^0$, the level-set mapping*

$$\omega \mapsto \{x \in \mathbb{R}^d \mid h(x,\omega) \leq \beta(\omega)\}$$

is measurable and closed-valued. A function $h : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a normal integrand if and only if

$$\operatorname{lev}_{\leq \beta} h(\omega) := \{x \in \mathbb{R}^n \mid h(x,\omega) \leq \beta\}$$

is measurable and closed-valued for every $\beta \in \mathbb{R}$.

Proof. Let $S(\omega) := \{x \in \mathbb{R}^d \mid h(x,\omega) \leq \beta(\omega)\}$. For a closed $C \subset \mathbb{R}^n$, $R(\omega) := C \times \{\alpha \mid \alpha \leq \beta(\omega)\}$ is measurable and closed-valued (an exercise). Now

$$\begin{aligned} S^{-1}(C) &= \{\omega \mid S(\omega) \cap C \neq \emptyset\} \\ &= \{\omega \mid S_h(\omega) \cap R(\omega) \neq \emptyset\} \\ &= \operatorname{dom}(\operatorname{epi} h \cap R) \end{aligned}$$

which shows the measurability of S.

To prove the second claim, note first that the closedness of the level sets implies that $\operatorname{epi} h$ is closed-valued. Let (β^ν) be a dense sequence in $\mathbb{R}$. Since countable unions of measurable mappings are measurable and

$$\operatorname{lev}_{<\beta^\nu} h(\omega) := \{x \in \mathbb{R}^n \mid h(x,\omega) < \beta^\nu\}$$

are measurable, we have

$$\operatorname{epi} h^o(\omega) = \bigcup_\nu \left(\operatorname{lev}_{<\beta^\nu} h(\omega) \times [\beta^\nu, \infty)\right),$$

so S_h is measurable. □

Theorem 3.8. *A normal integrand h is $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable from $\mathbb{R}^n \times \Omega$ to $\overline{\mathbb{R}}$. In particular,*

$$\omega \mapsto h(x(\omega), \omega)$$

is measurable for any $x \in L^0(\mathbb{R}^n)$.

Proof. Note that $\{[-\infty, \beta] \mid \beta \in \mathbb{R}\}$ generate the σ-algebra of $\overline{\mathbb{R}}$. For $\beta \in \mathbb{R}$, $\operatorname{lev}_{\leq \beta} h$ is closed-valued and measurable by Theorem 3.7, so $\{(x, \omega) \mid h(x, \omega) \leq \beta\}$ is measurable by Corollary 3.2. The second claim follows from the fact the compositions of measurable functions are measurable. □

Theorem 3.9. *The following are normal integrands:*

1. *$h(x, \omega) = \sup_{i \in \mathcal{J}} h^i(x, \omega)$, where h^i are normal integrands and $\mathcal{J}$ is countable.*
2. *$h(x, \omega) = \sum_{i=1}^n h^i(x, \omega)$, where h^i are normal integrands.*
3. *$h(\cdot, \omega) = \alpha(\omega) h^0(\cdot, \omega)$, where h^0 is a normal integrand. When $\alpha(\omega) = 0$, the scalar multiplication is defined as $\alpha(\omega) h^0(\cdot, \omega) = \operatorname{cl} \operatorname{dom} h^0(\cdot, \omega)$.*
4. *$h(x, \omega) = f(x, u(\omega), \omega)$, where $f : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ is a normal integrand and $u \in \mathcal{L}^0(\mathbb{R}^m)$ is measurable.*

Proof. Exercise (Hint: use Theorems 3.5, 3.6, 3.7 and 3.3).

In 1, $\operatorname{epi} h = \cap \operatorname{epi} h^i$, so the claim follows directly from Theorem 3.3. Below, in each case, M satisfies the assumption of Theorem 3.5, which can be verified using Theorems 3.6, 3.7 and 3.3.

To prove 2, let $S = \operatorname{epi}_{h^1} \times \cdots \times \operatorname{epi}_{h^J}$ and

$$M(x_1, \alpha_1, \ldots, x_J, \alpha_J) = \begin{cases} (x, \alpha_1 + \cdots + \alpha_J) & \text{if } x_1 = \cdots = x_J = x \\ \emptyset & \text{otherwise} \end{cases}$$

so that $\operatorname{epi} h = M \circ S$ and the claim follows from Theorem 3.5.1. The multiplication 3 follows similarly with

$$S(\omega) = \begin{cases} \operatorname{epi} h^0(\omega) & \text{if } \rho(\omega) > 0 \\ \operatorname{epi} \delta_{\operatorname{cl} \operatorname{dom} h(\cdot, \omega)} & \text{otherwise} \end{cases}$$

and $M(x, \alpha, \omega) = \{(x, \beta) \mid \rho(\omega)\alpha \leq \beta\}$. In 4, we have $\operatorname{epi} h(\omega) = \{x \mid M(x, \omega) \in \operatorname{epi} f(\cdot, \cdot, \omega)\}$ for $M(x, \alpha, \omega) := (x, u(\omega), \alpha)$, so Theorem 3.5.2. applies.

□

Theorem 3.10. *Assume that $f : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ is a normal integrand on $\mathbb{R}^n \times \mathbb{R}^m$ and let*

$$p(u, \omega) := \inf_{x \in \mathbb{R}^n} f(x, u, \omega).$$

The function defined by $\operatorname{cl}_u p(u, \omega)$ is a normal integrand on $\mathbb{R}^m$. In particular, if $p(\cdot, \omega)$ is lsc for every ω, p is a normal integrand.

Proof. Let $\Pi(x,u,\alpha) = (u,\alpha)$ be the projection from $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$ to $\mathbb{R}^m \times \mathbb{R}$. It is easy to check that $\Pi \operatorname{epi} f^o(\omega) = \operatorname{epi} p^o(\omega)$, so, for an open $O \subset \mathbb{R}^n \times \mathbb{R}$,

$$(\operatorname{epi} p^o)^{-1}(O) = (\operatorname{epi} f^o)^{-1}(\Pi^{-1}(O)),$$

where the right side is measurable, since f is a normal integrand and Π is continuous. Thus $\operatorname{epi} p$ is measurable. Moreover, $\omega \mapsto \operatorname{cl} \operatorname{epi} p$ is measurable, which shows that $(u,\omega) \mapsto (\operatorname{lsc}_u p)(u,\omega)$ is a normal integrand. If p is lsc in the first argument, it is thus a normal integrand. □

Corollary 3.11. *Given a normal integrand h, $p(\omega) := \inf_x h(x,\omega)$ is measurable, and*

$$S(\omega) := \operatorname*{argmin}_x h(x,\omega)$$

is measurable and closed-valued.

Proof. The first claim follows from Theorem 3.10. We have

$$S(\omega) = \{x \mid h(x,\omega) \le p(\omega)\},$$

so S is measurable by Theorem 3.7. Since $h(\cdot,\omega)$ is lsc, S is closed-valued. □

Example 3.12. *For a measurable closed-valued $S : \Omega \rightrightarrows \mathbb{R}^n$ and $x \in \mathcal{L}(\mathbb{R}^n)$, the projection mapping*

$$\omega \mapsto P_{S(\omega)}(x(\omega)) := \operatorname*{argmin}_{x' \in S(\omega)} |x' - x(\omega)|$$

is measurable and closed-valued. Indeed, this follows from Corollary 3.11 applied to $h(x',\omega) := \delta_{S(\omega)}(x') + |x' - x(\omega)|$.

Combining Theorem 3.10 and Lemma 2.12 gives the following.

Corollary 3.13. *Let h be a normal integrand such that $h(x) \ge -\rho|x| - m$ for $\rho, m \in \mathcal{L}^0_+$. The functions*

$$h^\nu(x,\omega) := \inf_{x' \in \mathbb{R}^d} \{h(x',\omega) + \nu\rho(\omega)|x - x'|\} \quad \nu \in \mathbb{N}$$

are Caratheodory integrands, $(\nu\rho)$-Lipschitz with $h^\nu(x) \ge -\rho|x| - m$ and as ν increases, they increase pointwise to h.

3.3 Measurable selections

A function $x : \Omega \to \mathbb{R}^d$ is called a *selection of S* if $x(\omega) \in S(\omega)$ for all $\omega \in \operatorname{dom} S$. The sequence (x^ν) of measurable selections of S in the following theorem is known as *a Castaing representation* of S.

Theorem 3.14 (Castaing representation). *Let $S : \Omega \rightrightarrows \mathbb{R}^d$ be closed-valued. Then S is measurable if and only if* $\operatorname{dom} S$ *is measurable and there exists a sequence $x^\nu \in \mathcal{L}^0(\mathbb{R}^n)$ such that, for all $\omega \in \operatorname{dom} S$,*

$$S(\omega) = \operatorname{cl}\{x^\nu(\omega) \mid \nu = 1, 2, \dots\}.$$

Proof. Assuming the Castaing representation exists, we have, for an open O,

$$S^{-1}(O) = \bigcup_{\nu=1}^{\infty} (x^\nu)^{-1}(O),$$

so S is measurable. Assume now that S is measurable. Let $\mathcal{J}$ be the countable collection of $q = (q_0, q_1, \dots, q_d)$, $q \in \mathbb{Q}^d$ such that $\{q_0, q_1, \dots, q_d\}$ are affinely independent. For each $q \in \mathcal{J}$, we define recursively $S^{q,0}(\omega) := P_{S(\omega)}(q_0)$ and

$$S^{q,i}(\omega) := P_{S^{q,i-1}(\omega)}(q_i).$$

These mappings are measurable and closed-valued, by Example 3.12. Moreover, $S^{q,d}(\omega)$ is a singleton, a point in $S(\omega)$ nearest to q_0. Setting $x^q(\omega) := S^{q,d}(\omega)$, (x^q) is a Castaing representation of S.

Let us verify that $S^{q,d}$ is single-valued. We fix ω and omit it from the notation. By the recursive definition of $S^{q,i}$, for each q^i, there is $r^i \geq 0$ such that $S^{q,d} \subset \partial\mathbb{B}(q^i, r^i)$. Thus, for any $x \in S^{q,d}$, $|x - q^i|^2 = (r^i)^2$ for all i. By affine independence, these equations have a unique solution. □

Corollary 3.15 (Measurable selection theorem). *Any measurable closed-valued $S : \Omega \rightrightarrows \mathbb{R}^n$ admits a measurable selection.*

Corollary 3.16 (Doob–Dynkin, set-valued version). *Let ξ be a random variable with values in a measurable space $(\Xi, \mathcal{A})$. A set-valued mapping S is a $\sigma(\xi)$-measurable closed random set if and only if there there exists measurable closed-valued $\tilde{S} : \Xi \rightrightarrows \mathbb{R}^n$ such that $S(\omega) = \tilde{S}(\xi(\omega))$. If S is convex-valued, $\tilde{S}$ can be chosen convex-valued.*

Proof. The sufficiency is clear. To prove necessity, let (x^ν) be a $\sigma(\xi)$-measurable Castaing representation of S. By Doob-Dynkin lemma **??**, there exist Borel-measurable $g^\nu : \Xi \to \mathbb{R}^n$ such that $x^\nu(\omega) = g^\nu(\xi(\omega))$. Let

$$\tilde{S}(y) = \operatorname{cl}\{g^\nu(y) \mid \nu = 1, 2, \dots\}$$

so that $S(\omega) = \operatorname{cl}\{g^\nu(\xi(\omega)) \mid \nu = 1, 2, \dots\} = \tilde{S}(\xi(\omega))$. If S is convex-valued, we can take closed convex hull of $\tilde{S}$, which is measurable by Corollary 3.4. □

Corollary 3.17 (Doob–Dynkin for normal integrands). *Let ξ be a random variable with values in a measurable space $(\Xi, \mathcal{A})$. A function h is a $\sigma(\xi)$-normal integrand on $\mathbb{R}^n$ if and only if there exists a $\mathcal{A}$-normal integrand H on $\mathbb{R}^n$ such that*

$$h(x, \omega) = H(x, \xi(\omega)).$$

If h is a convex normal integrand, H can be chosen a convex $\mathcal{A}$-normal integrand.

Proof. Apply Corollary 3.16 to the epigraphical mapping of h. ☐

Corollary 3.18. *A closed-valued mapping $S : \Omega \rightrightarrows \mathbb{R}^n$ is measurable if and only if there exists a sequence $x^\nu \in \mathcal{L}^0(\mathbb{R}^n)$ such that*

1. *for each ν, $\{\omega \mid x^\nu(\omega) \in S(\omega)\}$ is measurable,*
2. *for each ω, $S(\omega) \cap \{x^\nu(\omega) \mid \nu \in \mathbb{N}\}$ is dense in $S(\omega)$.*

Proof. Omitted. ☐

Theorem 3.19. *For normal integrands h and $\tilde{h}$, we have $h \leq \tilde{h}$ if and only if, for every $w \in \mathcal{L}^\infty(\mathbb{R}^n)$, $h(w) \leq \tilde{h}(w)$. In particular, $h = \tilde{h}$ if and only if $h(w) = \tilde{h}(w)$ for every $w \in \mathcal{L}^\infty(\mathbb{R}^n)$.*

Proof. Necessity is obvious. To prove the sufficiency, we may assume that $h \geq -m$ for some $m \in \mathbb{R}_+$. Indeed... ??. Let h^ν and $\tilde{h}^\nu$ be the respective Lipschitz-regularizations from Corollary 3.13. For a bounded measurable $w : \Omega \to \mathbb{R}^n$, we have $h(w) \leq \tilde{h}(w)$ if and only if $h^\nu(w) \leq \tilde{h}^\nu(w)$ for every ν. Thus it suffices to prove the claim for Lipschitz integrands h and $\tilde{h}$.

Assume for contradiction that, for some $\epsilon > 0$, the domain of

$$S(\omega) := \operatorname{lev}_{\leq -\epsilon}(\tilde{h} - h)(\cdot, \omega)$$

is nonempty. By Theorems 3.7 and 3.9, S is measurable. By Corollary 3.15, there is a measurable $\hat{w}$ with $\hat{w} \in S$ on $\operatorname{dom} S$. For ν large enough, $w := \hat{w}\mathbb{1}_{|\hat{w}| \leq \nu}$ and $A := \{|\hat{w}| \leq \nu\} \cap \operatorname{dom} S$ is nonempty which is a contradiction with the assumption that $h(w) \leq \tilde{h}(w)$. ☐

A function $M : \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ is a *Caratheodory mapping* if $M(\cdot, \omega)$ is continuous for all ω and $M(x, \cdot)$ is measurable for all $x \in \mathbb{R}^n$.

Theorem 3.20. *When M is a Caratheodory mapping,*

$$\operatorname{gph} M(\cdot, \omega) := \{(x, u) \mid u \in M(x, \omega)\}$$

defines a closed-valued measurable mapping.

Proof. For any countable dense $D \subset \mathbb{R}^n$, $\{(x, M(x, \omega)) \mid x \in D\}$ is a Castaing representation of $\operatorname{gph} M$. ☐

3.4 Convexity

Theorem 3.21. *Let $h : \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be such that, for almost every ω, the function $h(\cdot, \omega)$ is convex and its domain has nonempty interior. Then h is a convex normal integrand if and only if $h(x, \cdot)$ is measurable for every $x \in \mathbb{R}^n$.*

Proof. Omitted. □

For a convex-valued S, we define S^∞ as the ω-wise recession cone of S, i.e.

$$S^\infty(\omega) = \{x \in \mathbb{R}^n \mid \bar{x} + \lambda x \in S(\omega)\ \forall \bar{x} \in S(\omega),\ \lambda > 0\}.$$

If S is closed convex-valued and $\bar{x}(\omega) \in S(\omega)$, then, by Theorem **??** in the appendix,

$$S^\infty(\omega) = \bigcap_{\lambda>0} \lambda(S(\omega) - \bar{x}(\omega)),$$

so $S^\infty(\omega)$ is the largest closed convex cone that can be translated into $S(\omega)$.

Theorem 3.22. *If S is measurable and closed convex-valued, then so too is S^∞.*

Proof. By Corollary 3.15, there is $\bar{x} \in \mathcal{L}^0(\mathbb{R}^n)$ such that $\bar{x} \in S$ on $\operatorname{dom} S$. By convexity,

$$S^\infty(\omega) = \bigcap_{\nu=1}^{\infty} \frac{1}{\nu}(S(\omega) - \bar{x}(\omega))$$

for $\omega \in \operatorname{dom} S$. The measurability now follows from Theorem 3.3. □

Given a convex normal integrand h, we define h^∞ scenariowise as

$$h^\infty(\cdot, \omega) := h(\cdot, \omega)^\infty;$$

see the Appendix.

Theorem 3.23. *For a convex normal integrand h, h^∞ is a normal integrand. If h is proper, then*

$$h^\infty(x, \omega) = \sup_{\lambda>0} \frac{h(\bar{x}(\omega) + \lambda x) - h(\bar{x}(\omega), \omega)}{\lambda} \quad \forall (x, \omega) \in \mathbb{R}^n \times \Omega$$

for every $\bar{x} \in \mathcal{L}^0(\operatorname{dom} h)$.

Proof. By Theorem 3.22, h^∞ is a normal integrand. The formula follows from Theorem **??** in the appendix. □

Theorem 3.24. *Assume that f is a convex normal integrand and that the set-valued mapping*

$$N(\omega) = \{x \in \mathbb{R}^n \mid f^\infty(x, 0, \omega) \le 0\}$$

is linear-valued. Then

$$p(u, \omega) := \inf_{x \in \mathbb{R}^n} f(x, u, \omega)$$

is a normal integrand with

$$p^\infty(u, \omega) = \inf_{x \in \mathbb{R}^n} f^\infty(x, u, \omega).$$

Moreover, given a $u \in \mathcal{L}^0(\mathcal{F})$, there is an $x \in \mathcal{L}^0(\mathcal{F})$ with $x(\omega) \perp N(\omega)$ and

$$p(u(\omega), \omega) = f(x(\omega), u(\omega), \omega).$$

Proof. By Theorem 2.18, the linearity condition implies that the infimum in the definition of p is attained and that $p(\cdot,\omega)$ is a lower semicontinuous convex function with

$$p^\infty(u,\omega) = \inf_{x\in\mathbb{R}^n} f^\infty(x,u,\omega).$$

By Theorem 3.10, the lower semicontinuity implies that p is a normal integrand. By Theorem **??**, there is an $\bar{x}_t$ that attains the minimum for every ω. By Theorem 2.18, we may replace $\bar{x}_t(\omega)$ by its projection to the orthogonal complement of $N_t(\omega)$. By Example 3.12, the projection preserves measurability. □

Given an extended real-valued function g on $\mathbb{R}^m$ and an $\mathbb{R}^m$-valued function H on a subset $\operatorname{dom} H$ of $\mathbb{R}^n$, we define their composition as the extended real-valued function

$$(g\circ H)(x) := \begin{cases} g(H(x)) & \text{if } x \in \operatorname{dom} H,\\ +\infty & \text{if } x \notin \operatorname{dom} H.\end{cases}$$

Given a convex cone $K \subset \mathbb{R}^m$, the function H is said to be *K-convex* if the set

$$\operatorname*{epi}_K H := \{(x,u) \mid x \in \operatorname{dom} H,\ H(x) - u \in K\}$$

is convex. A K-convex function is *closed* if $\operatorname{epi}_K H$ is a closed set. It is easily verified (see the proof below) that if g is convex and H is K-convex then $h\circ H$ is convex if

$$H(x) - u \in K \implies g(H(x)) \le g(u) \quad \forall x \in \operatorname{dom} H. \tag{3.1}$$

We say that $H : \mathbb{R}^n\times\Omega \to \mathbb{R}^m$ is a *K-convex normal function* if $\omega \mapsto \operatorname{epi}_K H(\cdot,\omega)$ is closed convex-valued and measurable.

Theorem 3.25. *The following are convex normal integrands:*

1. *$h(\cdot,\omega) = \alpha(\omega)h^0(\cdot,\omega)$, where $\alpha \in \mathcal{L}^0_+$ and h^0 is a convex normal integrand. When $\alpha(\omega) = 0$, the scalar multiplication is defined as $\alpha(\omega)h^0(\cdot,\omega) = \operatorname{cl}\operatorname{dom} h^0(\cdot,\omega)$.*
2. *$h(x,\omega) = \sum_{i=1}^n h^i(x,\omega)$, where h^i are convex normal integrands.*
3. *$h = g\circ H$, where g is a convex normal integrand and H is a K-convex normal function such that (3.1) holds almost surely and $(-K) \cap \{u \in \mathbb{R}^m \mid g^\infty(u,\omega) \le 0\}$ is linear.*
4. *$h(x,\omega) = g(A(\omega)x,\omega)$ where g is a convex normal integrand and $A : \mathbb{R}^n \times \Omega \to \mathbb{R}^d$ is an affine Caratheodory mapping.*

Proof. The first two parts follow from Theorem 3.9 and the fact that scalar multiplication and the sum preserve convexity. By 2,

$$h(x,u,\omega) := g(u,\omega) + \delta_{\operatorname{epi}_K H(\cdot,\omega)}(x,u),$$

defines a convex normal integrand. The growth condition gives

$$(g \circ H)(x,\omega) = \inf_{u\in\mathbb{R}^m} h(x,u,\omega)$$

while the linearity condition implies, by Theorem 3.24, that this expression is a normal integrand. Part 4, follows from 3 by choosing $H = A$ and $K = \{0\}$. $\square$

Given a normal integrand h, we define h^* scenariowise, that is,

$$h^*(y,\omega) := \sup_u \{u\cdot y - h(u,\omega)\}.$$

Theorem 3.26. *Given a normal integrand h, h^* is a convex normal integrand.*

Proof. Let (x^ν,α^ν) be a Castaing representation of $\operatorname{epi} h$. On $\operatorname{dom}\operatorname{epi} h$,

$$h^*(y,\omega) = \begin{cases} \sup_\nu \{x^\nu(\omega)\cdot y - \alpha^\nu(\omega)\} & \omega \in \operatorname{dom}\operatorname{epi} h, \\ -\infty & \omega \notin \operatorname{dom}\operatorname{epi} h. \end{cases}$$

Being a countable supremum of normal integrands (in fact, Caratheodory integrands), h^* is normal. $\square$

In particular, for a measurable S,

$$\sigma_S(y,\omega) := \sigma_{S(\omega)}(v) := \sup_{x\in S(\omega)} x\cdot v$$

is a convex normal integrand. For a convex normal integrand h,

$$H(x,\alpha,\omega) := \begin{cases} \alpha h(x/\alpha,\omega) & \text{if } \alpha > 0, \\ h^\infty(x,\omega) & \text{if } \alpha = 0, \\ +\infty & \text{otherwise} \end{cases}$$

is a normal integrand. Indeed, it is easy to verify that H is the conjugate of the normal integrand defined by $\delta_{\operatorname{epi} h^*(\omega)}(v,-\beta)$. For a convex normal integrand h and $\alpha \in L^0(\mathbb{R}_+)$,

$$h^0(x,\omega) := \begin{cases} \alpha(\omega) h(x/\alpha(\omega),\omega) & \text{if } \alpha(\omega) > 0, \\ h^\infty(x,\omega) & \text{if } \alpha(\omega) = 0, \\ +\infty & \text{otherwise} \end{cases}$$

is a normal integrand. Indeed, this follows from $(h^0)^* = (\alpha h^*)^*$ and also from $h(x,\omega) = H(x,\alpha(\omega),\omega)$.

Theorem 3.27. *Given a convex normal integrand h, the set-valued mapping*

$$\omega \mapsto \operatorname{gph} \partial h(\cdot,\omega)$$

is closed-valued and measurable. In particular, given $x \in \mathcal{L}^0(\mathbb{R}^n)$, the mapping

$$\partial h(x) := \partial h(x(\omega),\omega)$$

is closed convex-valued and measurable.

Proof. We have

$$\operatorname{gph}\partial h(\cdot,\omega) = \{(x,v)\in\mathbb{R}^n\times\mathbb{R}^n \mid h(x,\omega)+h^*(v,\omega)-x\cdot v\le 0\},$$

so the closedness follows from the lower semicontinuity of h and h^* while the measurability follows from Theorems 3.27 and 3.7. The second claim now follows from Theorem 3.5 □

3.5 Integral functionals

Given a probability measure P on $(\Omega,\mathcal{F})$, $(\Omega,\mathcal{F},P)$ is a probability space. Let L^0 be the space of random variables, where random variables, that agree P-almost surely, are identified.

Given a normal integrand h, the associated integral functional $Eh : L^0 \to \overline{\mathbb{R}}$ is defined by

$$Eh(x) := \int h(x(\omega),\omega)dP(\omega).$$

Recall that $\omega \mapsto h(x(\omega),\omega)$ is measurable by Theorem 3.8. Here and in what follows, the expectation of an extended real-valued random variable is defined as $+\infty$ unless the positive part is integrable. Likewise, the sum of extended real numbers is defined as $+\infty$ if any of the numbers equals $+\infty$.

We say that two set-valued mappings S and $\bar{S}$ are indistinguishable if there exists a P-null set N such that $S=\bar{S}$ outside N. Normal integrands are said to indistinguishable if their epigraphical mappings are so. We say that a property holds for normal integrands on $\mathbb{R}^n$ *almost surely everywhere* if there exists a measurable set $\tilde{\Omega}\subset\Omega$ of full P measure such that the property holds on $\tilde{\Omega}\times\mathbb{R}^n$. In particular, normal integrands h and $\tilde{h}$ are indistinguishable if $h=\tilde{h}$ almost surely everywhere.

Theorem 3.28. *For normal integrands h and $\tilde{h}$, we have $h\le\tilde{h}$ almost surely everywhere if and only if $h(w)\le\tilde{h}(w)$ almost surely for every $w\in L^\infty(\mathbb{R}^n)$. In particular, $h=\tilde{h}$ almost surely everywhere if and only if $h(w)=\tilde{h}(w)$ almost surely for every $w\in L^\infty(\mathbb{R}^n)$. In this case, $h(w)=\tilde{h}(w)$ for every $w\in L^0(\mathbb{R}^n)$ and $Eh=E\tilde{h}$.*

Proof. The proof is analogous to that of Theorem 3.19. □

A vector space $\mathcal{X}\subseteq L^0$ is *decomposable* if $L^\infty\subset\mathcal{X}$ and $1_A x\in\mathcal{X}$ whenever $A\in\mathcal{F}$ and $x\in\mathcal{X}$. Equivalently, $\mathcal{X}$ is decomposable if

$$1_A x + 1_{\Omega\setminus A}x' \in \mathcal{U}$$

whenever $A\in\mathcal{F}$, $x\in\mathcal{X}$ and $x'\in L^\infty$. Examples of decomposable spaces include L^p-spaces with $p\ge 0$ and Orlicz spaces ??.

Theorem 3.29 (Interchange rule)**.** *For a normal integrand $h : \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and a decomposable $\mathcal{X}$, we have*

$$\inf_{x \in \mathcal{X}} Eh(x) = E[\inf_{x \in \mathbb{R}^n} h(x)] \tag{3.2}$$

if $\mathcal{X} \in L^0$ or the left side is less than $+\infty$. If this common value is finite,

$$\operatorname*{argmin}_{x \in \mathcal{X}} Eh = \{x \in \mathcal{X} \mid x \in \operatorname{argmin} h \ P\text{-a.s.}\}$$

Proof. By Theorem 3.10, the function p defined by

$$p(\omega) := \inf_{x \in \mathbb{R}^m} h(x, \omega)$$

is measurable. When $Ep = +\infty$, the claim is trivial. Let $\alpha > Ep$ and $\epsilon > 0$ be small enough so that $E\beta < \alpha$ for $\beta := \epsilon + \max\{p, -1/\epsilon\}$. The mapping

$$S(\omega) := \{x \mid h(x, \omega) \le \beta(\omega)\}$$

is measurable and closed-valued with $P(\operatorname{dom} S) = 1$, so, by Corollary 3.15, there exists $x \in L^0(\mathbb{R}^n)$ such that $h(x) \le \beta$ almost surely and thus $Eh(x) \le E\beta < \alpha$. When $\mathcal{X} = L^0$, this shows (3.2) since $\alpha > Ep$ was arbitrary. Let $\bar{x} \in \mathcal{X}$ be such that $Eh(\bar{x}) < \infty$, and define

$$x^\nu = 1_{|x| \le \nu} x + 1_{|x| > \nu} \bar{x}.$$

By construction, $x^\nu \in \mathcal{X}$, and $h(x^\nu) \le \max\{h(x), h(\bar{x})\}$ for all ν, so, by Fatou's lemma, $Eh(x^\nu) < E[h(x)] + \epsilon$ for ν large enough. Since $\alpha > Ep$ was arbitrary, this proves the first claim.

Assume now that Ep is finite. If $x' \in \operatorname{argmin} h$ almost surely, then $x' \in \operatorname{argmin} Eh$. If $x' \notin \operatorname{argmin} h$, then $P(A) > 0$ for $A := \{h(x') > \epsilon + p\}$ for some $\epsilon > 0$, and, as above, there is x such that $h(x) < h(x')$ on A, so $Eh(1_A x + 1_{A^C} x') < Eh(x')$, which means that $x' \notin \operatorname{argmin} Eh$. □

We equip L^0 with the translation invariant metric

$$d(x, x') := E\rho(|x' - x|),$$

where ρ is a bounded nondecreasing continuous function vanishing only at the origin. A sequence (x^ν) in L^0 *convergences in measure* to an $x \in L^0$ if

$$\lim_{\nu \to \infty} P(\{|x^\nu - x| \ge \epsilon\}) = 0$$

for all $\epsilon > 0$.

Lemma 3.30. *The space L^0 is a complete metric topological vector space where a sequence converges if and only if it converges in probability. A sequence converges in probability if and only if every subsequence has an almost surely convergent subsequence with a common limit.*

Proof. Let $(x^\nu)_{\nu\in\mathbb{N}}$ be a sequence in L^0 and $x \in L^0$. If $x^\nu \to x$ in probability, there is a subsequence $(x^{\nu_k})_{k\in\mathbb{N}}$ such that $P(\rho(|x^{\nu_k} - x|) \geq 2^{-k}) \leq 2^{-k}$. Let $A_k := \{\omega \mid \rho(|x^{\nu_k}(\omega) - x(\omega)|) \geq 2^{-k}\}$. By monotone convergence,

$$
\begin{aligned}
E[\sum_{k=1}^{\infty} \rho(|x^{\nu_k} - x|)] &= \sum_{k=1}^{\infty} E[\rho(|x^{\nu_k} - x|)] \\
&= \sum_{k=1}^{\infty} E[1_{A_k}\rho(|x^{\nu_k} - x|) + 1_{\Omega\setminus A_k}\rho(|x^{\nu_k} - x|)] \\
&\leq \sum_{k=1}^{\infty} E[1_{A_k} + 1_{\Omega\setminus A_k}2^{-k}] \\
&\leq \sum_{k=1}^{\infty}(2^{-k} + 2^{-k}) < \infty.
\end{aligned}
$$

Thus, $\sum_{k=1}^{\infty} \rho(|x^{\nu_k} - x|) < \infty$ almost surely so $\rho(|x^{\nu_k} - x|) \to 0$ almost surely. For the converse, assume that x^ν does not converge to x in probability. Then there is an $\epsilon > 0$ and a subsequence such that $P(|x^{\nu_k} - x| \geq \epsilon) > \epsilon$. By dominated convergence, this cannot hold for almost surely converging subsequences.

Let $x^\nu \to x$ in probability. By the first claim, every subsequence has an almost surely converging subsequence $x^{\nu_k} \to x$. By dominated convergence, $d(x^{\nu_k}, x) \to 0$. This implies that the whole sequence convergences in the L^0 metric. If x^ν does not converge to x in probability, there is an $\epsilon > 0$, $\delta > 0$ and a subsequence such that $P(|x^{\nu'} - x| \geq \epsilon) \geq \delta$. Then $d(x^{\nu'}, x) \geq \delta\rho(\epsilon)$, so subsequences of $(x^{\nu'})$ cannot converge to x in L^0.

If $(x^\nu)_{\nu\in\mathbb{N}}$ is Cauchy in L^0, there is a subsequence $(x^{\nu_k})_{k\in\mathbb{N}}$ such that $d(x^{\nu_{k+1}}, x^{\nu_k}) \leq 2^{-k}$. By monotone convergence,

$$
E[\sum_{k=1}^{\infty} \rho(|x^{\nu_{k+1}} - x^{\nu_k}|)] = \sum_{k=1}^{\infty} E[\rho(|x^{\nu_{k+1}} - x^{\nu_k}|)] \leq \sum_{k=1}^{\infty} 2^{-k} < \infty,
$$

so $\sum_{k=1}^{\infty} \rho(|x^{\nu_{k+1}} - x^{\nu_k}|) < \infty$ almost surely. Thus, $(x^{\nu_k})_{k\in\mathbb{N}}$ is almost surely Cauchy in $\mathbb{R}^n$ so it converges almost surely to an $x \in L^0$. By dominated convergence, $x^{\nu_k} \to x$ in L^0. The triangle inequality now implies that the whole sequence converges to x. □

Exercise 3.5.1. *For $\Omega = [0, 1]$ and the Lebesque measure P, show that L^0 is not locally convex. Hint: show that any nonempty open convex set is the whole L^0.*

We say that a normal integrand h is *bounded from below* if there is an $m \in L^1$ such that

$$
h(x, \omega) \geq m(\omega) \quad \forall x \in \mathbb{R}^n,\ \omega \in A,
$$

where $A \in \mathcal{F}$ is of full measure.

Theorem 3.31. *If h is a normal integrand bounded from below, then Eh is lsc on L^0.*

Proof. Let $x^\nu \to x$ in L^0. By Lemma 3.30, we may assume, by passing to a subsequence if necessary, that $x^\nu \to x$ almost surely. By lower semicontinuity of $h(\cdot,\omega)$,

$$\liminf_{\nu\to\infty} h(x^\nu(\omega),\omega) \geq h(x(\omega),\omega)$$

almost surely so lower semicontinuity follows from Fatou's lemma. □

Remark 3.32. *The lower boundedness assumption in Theorem 3.31 can be relaxed as follows. Recall that A is an atom of P if for every $B \subset A$, $P(B) = 0$ or $P(B) = P(A)$.*

Given a normal integrand h such that Eh is proper on L^0, Eh is lsc on L^0 if and only if

$$A := \{\omega \in \Omega \mid \inf_x h(x,\omega) = -\infty\}$$

contains only atoms of P and at most finitely many of them, and there exists $m \in L^1$ such that $h \geq -m$ on A^C. If Eh is not lsc, $\operatorname{lsc} Eh = -\infty$ on $\operatorname{dom} Eh$.

The proof is left as an exercise.

Lemma 3.33. *Given extended real-valued random variables ξ_1 and ξ_2, we have*

$$E[\xi_1 + \xi_2] = E[\xi_1] + E[\xi_2]$$

under any of the following:

1. $\xi_1^+, \xi_2^+ \in L^1$,
2. $\xi_1 \in L^1$ *or* $\xi_2 \in L^1$,
3. $\xi_1^-, \xi_2^- \in L^1$.
4. ξ_1 *or* ξ_2 *is* $\{0, +\infty\}$*-valued.*

Proof. Exercise. □

The following is an immediate corollary of Lemma 3.33.

Lemma 3.34. *Given normal integrands h_1 and h_2, we have*

$$E[h_1 + h_2] = E[h_1] + E[h_2]$$

under any of the following:

1. *the integrands are lower bounded.*
2. *h_1 or h_2 is an indicator function of a measurable closed-valued mapping.*

If h is a convex normal integrand, then Eh is a convex function on L^0.

Theorem 3.35. *If h is a convex normal integrand such that Eh is lsc and proper, then*

$$(Eh)^\infty = Eh^\infty.$$

Proof. Let $\bar{x} \in \operatorname{dom} Eh$. Since Eh is lsc and $h(\bar{x})$ is integrable, Theorem **??** and Lemma 3.33 give

$$(Eh)^\infty(x) = \sup_{\lambda>0} E\left[\frac{h(\lambda x + \bar{x}) - h(\bar{x})}{\lambda}\right].$$

The difference quotients

$$h^\lambda(x(\omega),\omega) := \frac{h(\lambda x(\omega) + \bar{x}(\omega),\omega) - h(\bar{x}(\omega),\omega)}{\lambda}$$

increase pointwise to $h^\infty(x(\omega),\omega)$. Thus, $(Eh)^\infty \le Eh^\infty$. If $x + \bar{x} \notin \operatorname{dom} Eh$, then $(Eh)^\infty(x) = +\infty$. If $x + \bar{x} \in \operatorname{dom} Eh$, the claim follows from the monotone convergence theorem. □

3.6 Existence of solutions

The sequence x^{μ^ν} in the next lemma is called a *random subsequence*.

Lemma 3.36. *For an almost surely bounded sequence (x^ν) in $\mathcal{N}$, there exists $\mathcal{F}_T$-measurable integer-valued functions (μ^ν) and $x \in \mathcal{N}$ such that $x^{\mu^\nu} \to x$ almost surely.*

Proof. We will prove first that, for an almost surely bounded sequence (η_ν) in $L^0(\mathcal{G};\mathbb{R}^d)$, there exists $\mathcal{G}$-measurable integer-valued functions (μ^ν) and $\eta \in L^0(\mathcal{G};\mathbb{R}^d)$ such that $\eta_{\mu^\nu} \to \eta$ almost surely.

Let $\bar{\eta}^1 = \limsup_\nu \eta^1_\nu$, $\mu^1_0 = 0$ and $\mu^1_{\nu+1} = \inf\{\nu' > \nu \mid |\eta^1_{\nu'} - \bar{\eta}^1| \le 1/\nu\}$ so that $\eta^1_{\mu^1_\nu} \to \bar{\eta}^1$. Applying this to such iteratively constructed $(\eta_{\mu^i_\nu})$ iteratively to each component, we arrive at an sequence (μ^d_ν) such that $\eta_{\mu^d_\nu} \to \bar{\eta}$ almost surely for some $\bar{\eta} \in L^0(\mathcal{G};\mathbb{R}^d)$.

Applying the above to $(x^\nu_0)_{\nu=1}^\infty$ we get an $\mathcal{F}_0$-measurable random subsequence μ^ν_0 such that $x_0^{\mu^\nu_0} \to x_0$ for an $x_0 \in L^0(\Omega,\mathcal{F}_0,P;\mathbb{R}^{n_0})$. Applying the above next to $(x_1^{\mu^\nu_0})_{\nu=1}^\infty$ we get an $\mathcal{F}_1$-measurable subsequence μ^ν_1 of μ^ν_0 such that $x_1^{\mu^\nu_1} \to x_1$ for an $x_1 \in L^0(\Omega,\mathcal{F}_1,P;\mathbb{R}^{n_1})$. Since $x_0^{\mu^\nu_0} \to x_0$ we also have $x_0^{\mu^\nu_1} \to x_0$. Extracting further subsequences similarly for $t = 2,\dots,T$ we arrive at the conclusion. □

Recall that, given a closed convex-valued mapping C, the closed convex-valued mapping C^∞, defined scenariowise as $C^\infty(\omega) := C(\omega)^\infty$, is measurable by Theorem 3.22.

Lemma 3.37. *Let C be a closed convex-valued mapping. Then every sequence in $\mathcal{N}(C)$ is almost surely bounded if and only if $\mathcal{N}(C^\infty) = \{0\}$.*

Proof. If $\{x \in \mathcal{N} \mid x \in C^\infty \text{ a.s.}\} \neq \{0\}$, then $\{x \in \mathcal{N} \mid x \in C \text{ a.s.}\}$ contains a half-line and thus an unbounded sequence. Assume now that

$$\{x \in \mathcal{N} \mid x \in C^\infty \text{ a.s.}\} = \{0\}$$

and let (x^ν) be a sequence in $\mathcal{N}(C)$ By a translation with an adapted process, we may assume that $0 \in C$ almost surely. Indeed, the translation does not affect any of the conditions of the statement. Assume that the claim holds for any $T-1$-period model.

If $\rho := \sup |x_0^\nu| < \infty$ almost surely, let

$$\begin{aligned} \mathcal{N}_1 &:= \{(x_1, \dots, x_T) \mid x_t \in L^0(\mathcal{F}_t)\} \\ C_1(\omega) &:= \{(x_1, \dots, x_T) \mid \exists x_0 \in \rho(\omega)\mathbb{B} : \ (x_0, \dots, x_T) \in C(\omega)\}, \end{aligned}$$

so that the results in Sections 5 and 6 give that C_1 is a measurable closed-convex mapping with

$$C_1^\infty(\omega) = \{(x_1, \dots x_T) \mid (0, x_1, \dots, x_T) \in C^\infty(\omega)\}$$

and hence the induction hypotheses gives that $(x_1^\nu, \dots, x_T^\nu)$ is bounded since $\mathcal{N}_1(C_1^\infty) = \{0\}$.

Assume now that $A(\omega) = \{\sup x_0^\nu = \infty\}$ has positive probability. Let $\alpha^\nu = 1_A/(|x_0^\nu| \vee 1)$ and $\bar{x}^\nu = \alpha^\nu x^\nu$. Passing to a random subsequence, we may assume that $\alpha^\nu \searrow 0$ almost surely. We have $\bar{x}^\nu \in \mathcal{N}$, $\bar{x}^\nu \in \alpha^\nu C$ and $|\bar{x}_0^\nu| \leq 1$. Since $\alpha^\nu \leq 1$, $\alpha^\nu C \subset C$ by convexity. By the previous paragraph, $(\bar{x}^\nu)$ is almost surely bounded and thus there is a random subsequence (τ^ν) such that $\bar{x}^{\tau^\nu} \to \bar{x} \in \mathcal{N}$ almost surely. By Exercise 2.5.1, $\bar{x} \in C^\infty$, so $\bar{x} = 0$ by assumption. This is a contradiction, since $|\bar{x}_0| = 1$ on A by construction.

To start the induction for $T = 0$, the argument is the same as in the previous paragraph except we do not need to refer to the earlier paragraph. □

Theorem 3.38 (Komlós)**.** *Let $(x^\nu)_{\nu \in \mathbb{N}}$ be a sequence in $L^0(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ that satisfies one of the following conditions:*

1. *$(x^\nu)_{\nu \in \mathbb{N}}$ is almost surely bounded in the sense that*

$$\sup_{\nu \in \mathbb{N}} |x^\nu(\omega)| < \infty \quad a.s.;$$

2. *$(x^\nu)_{\nu \in \mathbb{N}}$ is bounded in L^1.*

Then there exists a sequence $(\bar{x}^\nu)_{\nu \in \mathbb{N}}$ of convex combinations $\bar{x}^\nu \in \operatorname{co}\{x^\mu \mid \mu \geq \nu\}$ that converges almost surely to an $\mathbb{R}^n$-valued random variable.

Proof. Omitted. □

The following gives sufficient conditions for existence of solutions. The conditions will be generalized first in Theorem 3.43 below and later in Chapter **??** after the development of the dynamic programming principle.

Theorem 3.39. *Assume that h is a lower bounded convex normal integrand such that*

$$\{x \in \mathcal{N} \mid h^\infty(x) \le 0\} = \{0\}$$

almost surely. Then (SP) has an optimal solution.

Proof. Let $(x^\nu) \in \mathcal{N}$ be such that $Eh(x^\nu) \to \inf(SP)$. There exists $\gamma \in \mathbb{R}$ such that

$$Eh(x^\nu) \le \gamma.$$

Komlos theorem (Lemma 3.38) gives a sequence of convex combinations

$$\phi^\nu(\omega) := \sum_{\mu=\nu}^{\infty} \alpha^{\nu,\mu} h(x^\mu(\omega), \omega)$$

that converges almost surely to a real-valued measurable function. In particular, the function $\phi(\omega) := \sup_\nu \phi^\nu(\omega)$ is almost surely finite. Defining

$$\bar{x}^\nu = \sum_{\mu=\nu}^{\infty} \alpha^{\nu,\mu} x^\mu$$

we have by convexity that

$$h(\bar{x}^\nu(\omega), \omega) \le \phi^\nu(\omega) \le \phi(\omega) \quad P\text{-a.s.}$$

and $Eh(\bar{x}^\nu) \to \inf(SP)$. Then each $\bar{x}^\nu$ is a selection of

$$C(\omega) = \{x \mid h(x, \omega) \le \phi(\omega)\}.$$

Theorem 2.17 gives

$$C^\infty(\omega) = \{x \in \mathbb{R}^n \mid h^\infty(x, \omega) \le 0\},$$

so $(\bar{x}^\nu)$ is almost surely bounded by Lemma 3.37.

By Lemma 3.38, there is a sequence $(\hat{x}^\nu)_{\nu=1}^\infty$ of convex combinations of $(\bar{x}^\nu)_{\nu=1}^\infty$ that converges almost surely to a point x. By convexity, $Ek(\hat{x}^\nu) \to \inf(SP)$. The function Ek is lsc on L^0 by Theorem 3.31, so

$$Eh(x) \le \liminf_{\nu \to \infty} Eh(\hat{x}^\nu) = \inf(SP),$$

which completes the proof. □

Clearly, the second condition in Theorem 3.39 holds if, for P-almost every $\omega \in \Omega$,

$$\{x \in \mathbb{R}^n \mid h^\infty(x,\omega) \le 0\} = \{0\},$$

which means that $h(\cdot,\omega)$ is inf-compact. In the deterministic setting, the condition simply means that the level sets of h are bounded.

The lower boundedness in Theorem 3.39 can be relaxed significantly using the following very useful result. We denote

$$\mathcal{N}^\perp := \{v \in L^1(\mathbb{R}^n) \mid E[x \cdot v] = 0 \ \forall x \in \mathcal{N}^\infty\},$$

where $\mathcal{N}^\infty := \mathcal{N} \cap L^\infty$.

Lemma 3.40. *Let $x \in \mathcal{N}$ and $v \in \mathcal{N}^\perp$. If $E[x \cdot v]^+ \in L^1$, then $E[x \cdot v] = 0$.*

Proof. Assume first that $T = 0$. Defining $x^\nu := \mathbb{1}_{\{|x| \le \nu\}} x$, we have $x^\nu \in L^\infty$, so $E[x^\nu \cdot v] = 0$ and

$$E[x \cdot v]^- \le \liminf_{\nu \to \infty} E[x^\nu \cdot v]^- = \liminf_{\nu \to \infty} E[x^\nu \cdot v]^+ \le E[x \cdot v]^+,$$

where the inequalities follow from Fatou's lemma. By dominated convergence, $E[x \cdot v] = \lim E[x^\nu \cdot v] = 0$.

Assume now that the claim holds for every $(T-1)$-period model. Defining $x^\nu := \mathbb{1}_{\{|x_0| \le \nu\}} x$, we have

$$[\sum_{t=1}^T x_t^\nu \cdot v_t]^+ \le [x^\nu \cdot v]^+ + [x_0^\nu \cdot v_0]^- \le [x \cdot v]^+ + [x_0^\nu \cdot v_0]^-,$$

so $E[\sum_{t=1}^T x_t^\nu \cdot v_t] = 0$, by the induction hypothesis. Since $x_0^\nu \in L^\infty$, we get $E[x^\nu \cdot v] = 0$. It then follows that $E[x \cdot v] = 0$ just like in the case $T = 0$. ☐

Example 3.41. *If a martingale s and $x \in \mathcal{N}$ are such that*

$$E[\sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}]^+ < \infty,$$

then $E[\sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}] = 0$. This follows from Lemma 3.40 with $v \in \mathcal{N}^\perp$ defined by $v_t = \Delta s_{t+1}$.

Definition 3.42. *A normal integrand h is $\mathcal{N}_p^\perp$-bounded if there exists $p \in \mathcal{N}^\perp$ and $m \in L^1$ such that*

$$h(x,\omega) \ge x \cdot p(\omega) - m(\omega),$$

and

$$Eh(x) = E[h(x) - x \cdot p] \quad \forall x \in \mathcal{N}.$$

Theorem 3.43. *Assume that h is an $\mathcal{N}_p^\perp$-bounded convex normal integrand such that*

$$\{x \in \mathcal{N} \mid h^\infty(x) - x \cdot p \le 0\} = \{0\}$$

almost surely. Then (SP) has an optimal solution.

Proof. Let $k(x,\omega) := h(x,\omega) - x \cdot p(\omega)$. By assumption, $Ek = Eh$ on $\mathcal{N}$, so Theorem 3.39 proves the claim. □

Corollary 3.44. *If h is an $\mathcal{N}_p^\perp$-bounded normal integrand with*

$$p \in \operatorname{int}\operatorname{dom} h^* \quad P\text{-}a.s.,$$

then (SP) has optimal solutions.

Proof. By Theorem **??**, the interiority condition means that $h^\infty(x) > x \cdot p$ for all $x \in \mathbb{R}^n \setminus \{0\}$, or in other words,

$$\{x \in \mathbb{R}^n \mid h^\infty(x) - x \cdot p \le 0\} = \{0\} \quad P\text{-a.s.}$$

Thus the claim follows from Theorem 3.39. □

Lemma 3.45. *A normal integrand h is $\mathcal{N}_p^\perp$-bounded if there exists $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ with*

$$\lambda p \in \operatorname{dom} Eh^*$$

for all $\lambda \in [1-\epsilon, 1+\epsilon]$. In this case

$$\{x \in \mathcal{N} \mid h^\infty(x) \le 0 \text{ a.s.}\} = \{x \in \mathcal{N} \mid h^\infty(x) - x \cdot p \le 0 \text{ a.s.}\}.$$

Proof. The assumption implies $Eh^*(p) < \infty$ and $Eh^*((1+\epsilon)p) < \infty$. By Fenchel's inequality,

$$h(x,\omega) \ge x \cdot p(\omega) - h^*(p(\omega),\omega),$$
$$h(x,\omega) - x \cdot p(\omega) \ge \epsilon x \cdot p(\omega) - h^*((1+\epsilon)p(\omega),\omega).$$

Let $x \in \mathcal{N}$. If either $Eh(x) < \infty$ or $E[h(x) - x \cdot p] < \infty$, the above inequalities and Lemma 3.40 give $E[x \cdot p] = 0$, so

$$Eh(x) = E[h(x) - x \cdot p].$$

The above inequalities also give

$$h^\infty(x,\omega) \ge x \cdot p(\omega),$$
$$h^\infty(x,\omega) - x \cdot p(\omega) \ge \epsilon x \cdot p(\omega).$$

If either $h^\infty(x,0) \le 0$ or $h^\infty(x,0) - x \cdot p \le 0$ almost surely, then $x \cdot p \le 0$ almost surely. Lemma 3.40 then implies $x \cdot p = 0$ almost surely, which proves the last claim. □

Note that $\lambda p \in \operatorname{dom} Eh^*$ means that

$$h(x,\omega) \geq \lambda x \cdot p(\omega) - m(\omega)$$

for some $m \in L^1$. Thus, h satisfies the assumptions of Lemma 3.45 if and only if there exists $p \in \mathcal{N}^\perp$, $\epsilon > 0$ and $m \in L^1$ such that

$$h(x,\omega) \geq x \cdot p(\omega) + \epsilon|x \cdot p(\omega)| - m(\omega).$$

In particular, the condition holds when h is bounded from below by an integrable random variable. In the deterministic setting, the condition simply means that h is bounded from below. More interesting examples will be given in Section 3.7.

3.7 Applications

Exercise 3.7.1. *Verify that h in each application below is indeed a normal integrand.*

3.7.1 Mathematical programming

Example 3.46 (Mathematical programming)**.** *Consider the problem*

$$\begin{array}{ll} \text{minimize} & Eh_0(x) \quad \text{over } x \in \mathcal{N} \\ \text{subject to} & h_j(x) \leq 0 \quad P\text{-}a.s.,\ j = 1,\dots,m, \end{array}$$

where h_j are normal integrands. If there is a $p \in \mathcal{N}^\perp$, $\epsilon > 0$ and an $m \in L^1$ such that

$$h_0(x) \geq x \cdot p + \epsilon|x \cdot p| - m \quad P\text{-}a.s.$$

for all $x \in \mathbb{R}^n$ with

$$h_j(x) \leq 0 \quad j = 1,\dots,m \quad P\text{-}a.s.$$

then the problem has a solution as soon as

$$\{x \in \mathcal{N} \mid h_j^\infty(x) \leq 0\ P\text{-}a.s.\ \forall j = 0,\dots,m\} = \{0\}.$$

Proof. This fits the general format of (SP) with

$$h(x,\omega) = \begin{cases} h_0(x,\omega) & \text{if } h_j(x,\omega) \leq 0 \text{ for } j = 1,\dots,m, \\ +\infty & \text{otherwise.} \end{cases}$$

Indeed, by Theorem **??**, h is a normal integrand and

$$Eh(x) = \begin{cases} Eh_0(x) & \text{if } h_j(x) \leq 0\ P\text{-a.s. } j = 1,\dots,m, \\ +\infty & \text{otherwise.} \end{cases}$$

By ??,

$$h^\infty(x,\omega)=\begin{cases}h_0^\infty(x,\omega) & \text{if } h_j^\infty(x,\omega)\le 0 \text{ for } j=1,\dots,m,\\ +\infty & \text{otherwise,}\end{cases}$$

so the claim follows from Theorem 3.39 □

Example 3.47 (Composite models). *The above formats can be extended to*

$$h(x,\omega)=h_0(x,\omega)+g(H(x,\omega),\omega),$$

where H is a random K-convex function from $\mathbb{R}^n$ to $\mathbb{R}^m$ and g is a convex normal integrand on $\mathbb{R}^m$ satisfying (3.1)*. Choosing $g=\delta_{\mathbb{R}^m_-}$ we recover Example 3.46.*

In the linear case, Example 3.46 can be written as follows.

Example 3.48 (Linear programming). *Consider the problem*

$$\begin{array}{ll}\text{minimize} & E[c\cdot x] \quad \text{over } x\in\mathcal{N}\\ \text{subject to} & Ax\le b \quad P\text{-a.s.},\end{array}$$

Assume that there exists $p\in\mathcal{N}^\perp$ and $\epsilon>0$ such that

$$E\inf_{x\in\mathbb{R}^n}\{x\cdot(c-\lambda p)\mid Ax\le b\}>-\infty$$

for $\lambda\in[1-\epsilon,1+\epsilon]$. The problem has a solution as soon as

$$\{x\in\mathcal{N}\mid c\cdot x\le 0,\ Ax\le 0\ P\text{-a.s.}\}=\{0\}.$$

3.7.2 Stochastic control and problems of Bolza

Example 3.49 (Stochastic control). *Consider the problem*

$$\begin{array}{ll}\text{minimize} & E\left[\sum_{t=0}^T L_t(X_t,U_t)\right] \quad \text{over } (X,U)\in\mathcal{N},\\ \text{subject to} & \Delta X_t=A_tX_{t-1}+B_tU_{t-1}+u_t \quad t=1,\dots,T,\end{array}$$

where X and U are processes of fixed dimension, A_t and B_t are $\mathcal{F}_t$-measurable random matrices and u_t is an $\mathcal{F}_t$-measurable random vector all of appropriate dimensions. The functions L_t are $\mathcal{F}_t$-measurable convex normal integrands.

This is a classical formulation of convex stochastic optimal control where X describes the state of the controlled system and U is the control. If all L_t are bounded from below and if L_0 and $L_t(x,\cdot)$ for $t=1,\dots,T$ are inf-compact for all x, then an optimal solution exists.

Proof. This fits the general framework with $x = (X, U)$,

$$h(x) = \sum_{t=0}^{T} L_t(X_t, U_t) + \sum_{t=1}^{T} \delta_{\{0\}}(\Delta X_t - A_t X_{t-1} - B_t U_{t-1} - u_t),$$

Indeed, by Theorem **??**, h is a convex normal integrand and by ??,

$$h^\infty(x) = \sum_{t=0}^{T} L_t^\infty(X_t, U_t) + \sum_{t=1}^{T} \delta_{\{0\}}(\Delta X_t - A_t X_{t-1} - B_t U_{t-1}),$$

so the claim follows from Theorem 3.39. □

Example 3.50 (Stochastic problem of Bolza)**.** *Consider the problem*

$$\text{minimize} \quad E[\sum_{t=0}^{T} K_t(x_{t-1}, \Delta x_t) + k(x_T)] \quad \text{over } x \in \mathcal{N}, \tag{3.3}$$

where x is a process of fixed dimension d, k is a normal integrand, K_t are $\mathcal{F}_t$-measurable convex normal integrands and $x_- = 0$. Assume that

1. *There exists $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that for any $\lambda \in [1-\epsilon, 1+\epsilon]$, there are $y \in \mathcal{N}^1$ and $m_t \in L^1$ with*

$$\begin{aligned} K_t(x_{t-1}, \Delta x_t) &\geq \lambda((p_{t-1} - y_{t-1}) \cdot x_{t-1} + y_t \cdot x_t) - m_t, \\ k(x_T, \omega) &\geq -\lambda y_T \cdot x_T - m_{T+1}. \end{aligned}$$

Optimal solutions exist as soon as

$$\{x \in \mathcal{N} \mid \sum_{t=0}^{T} K_t^\infty(x_{t-1}, \Delta x_t) + k^\infty(x_T) \leq 0 \text{ a.s.}\} = \{0\}.$$

In particular, If $K_t(x, \cdot)$ are inf-compact for all x and $t = 0, \dots, T$, then an optimal solution exists.

Proof. This fits the general format with

$$h(x, \omega) = \sum_{t=0}^{T} K_t(x_{t-1}, \Delta x_t, \omega) + k(x_T, \omega),$$

so the claim follows from Theorem 3.39. □

3.7.3 Financial mathematics

Later, we formulate optimal investment problem as stochastic control.

Example 3.51 (Financial mathematics). *Let $s = (s_t)_{t=0}^T$ be an adapted $\mathbb{R}^J$-valued stochastic process describing the unit prices or assets in a perfectly liquid financial market. Consider the problem of finding a dynamic trading strategy $z = (z_t)_{t=0}^T$ that provides the "best hedge" against a financial liability of delivering a random amount $c \in L^0$ cash at time T. If we measure our risk preferences over random cash-flows with the "expected shortfall" associated with a nondecreasing nonconstant convex "loss function" $V : \mathbb{R} \to \overline{\mathbb{R}}$, the problem can be written as*

$$\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}_D, \tag{3.4}$$

where $\mathcal{N}_D$ denotes the set of adapted trading strategies $z = (z_t)_{t=0}^T$ that satisfy the portfolio constraints $z_t \in D_t$ for all $t = 0, \dots, T$ almost surely. Here D_t is a random $\mathcal{F}_t$-measurable set consisting of the portfolios we are allowed to hold over time period $(t, t+1]$.

The problem admits a solution if there exists a P-absolutely continuous martingale measure Q of the price process s such that, for $y := dQ/dP$, $yu \in L^1$ and $EV^*(\lambda y) < \infty$ for $\lambda \in [1-\epsilon, 1+\epsilon]$ and if

$$\{x \in \mathcal{N} \mid \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1} \geq 0,\ z_t \in D_t^\infty\} = \{0\}.$$

This last condition says that the only completely riskless strategy is the one that does not invest in the risky assets.

Proof. The problem fits the general framework with

$$h(x,\omega) = V\left(u(\omega) - \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1}(\omega)\right) + \sum_{t=0}^{T-1} \delta_{D_t}(z_t, \omega).$$

Indeed, h is a convex normal integrand by Theorem **??**. We have that h is $\mathcal{N}_p^\perp$-bounded with $p_t := -y\Delta s_{t+1}$. Indeed, Fenchel's inequality gives

$$h(x) \geq -\lambda^i y \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1} - \lambda^i yu - V^*(\lambda^i y),$$

and

$$m = \max\{\lambda^1 yu + V^*(\lambda^1 y), \lambda^2 yu + V^*(\lambda^2 y)\}.$$

Since V is a nonconstant function, we have $V^\infty(u) > 0$ for $u > 0$ and hence

$$V^\infty(-\sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}) \leq 0 \quad \Leftrightarrow \quad \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \geq 0$$

so

$$\{x \in \mathbb{R}^n \mid h^\infty(x,\omega) \le 0\} = \{x \in \mathbb{R}^n \mid \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega) \ge 0,\, x_t \in D_t(\omega)\}$$

and the recession condition in Theorem 3.39 becomes the last condition in the statement. □

Example 3.52 (Semistatic hedging). *Consider the problem*

$$\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1} - c \cdot \bar{z}_0 + S_0(\bar{x}_0)\right) \quad \text{over} \quad z \in \mathcal{N}_D, \bar{z}_0 \in \mathbb{R}^J,$$

Example 3.53 (Currency markets). *Consider Example 3.50 with*

$$K_t(x_{t-1}, \Delta x_t, \omega) = \delta_{D_{t-1}}(x_{t-1}, \omega) + \delta_{C_t}(\Delta x_t, \omega)$$

for adapted sequences $(D_t)_{t=0}^T$ and $(C_t)_{t=0}^T$ of closed convex random sets. This model can be used to describe trading in currency markets. Indeed, the $D_t(\omega)$ can be used to describe portfolio constraints while $C_t(\omega)$ models portfolios that are freely available in the market.

4 Integral functionals in duality

Convex duality is based on the theory of conjugate functions on dual pairs of locally convex topological vector spaces; see Sections **??** and 2.9. The first part of this section reviews dual pairs of spaces of random variables while the second part reviews conjugation of integral functionals on such spaces. This forms the functional analytic setting for the duality theory of stochastic optimization developed in the following sections. For full generality, we make minimal assumptions on the spaces of random variables.

4.1 Dual spaces of random variables

Let $\mathcal{U}$ and $\mathcal{Y}$ be decomposable linear spaces (see Section **??**) of $\mathbb{R}^m$-valued random variables such that $u \cdot y \in L^1$ for all $u \in \mathcal{U}$ and $y \in \mathcal{Y}$. We will assume that $\mathcal{U}$ and $\mathcal{Y}$ are in *separating duality* under the bilinear form

$$\langle u, y\rangle := E[u \cdot y]$$

in the sense that for every nonzero $u \in \mathcal{U}$, there exists a $y \in \mathcal{Y}$ such that $\langle u, y\rangle \neq 0$ and vice versa. As usual, we identify random variables that coincide almost surely so the elements of $\mathcal{U}$ and $\mathcal{Y}$ are actually *equivalence classes* of random variables.

We will also assume that the spaces are *solid* in the sense that if $\bar{u} \in \mathcal{U}$ and $u \in L^0$ are such that $|u^i| \leq |\bar{u}^i|$ almost surely for every $i = 1, \dots, m$, then $u \in \mathcal{U}$; similarly for $\mathcal{Y}$. Solidity implies that

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_m \quad \text{and} \quad \mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m, \tag{4.1}$$

where $\mathcal{U}_i$ and $\mathcal{Y}_i$ are solid decomposable linear spaces of real-valued random variables in separating duality under the bilinear form $(u_i, y_i) \mapsto E[u_i y_i]$. In particular,

$$u_i y_i \in L^1 \quad \text{and} \quad \langle u, y \rangle = \sum_{i=1}^m E[u_i y_i] \quad \forall u \in \mathcal{U},\ y \in \mathcal{Y}. \tag{4.2}$$

Remark 4.1. *A linear space $\mathcal{U}$ of random variables is solid, in particular, if*

$$u \in L^0,\ \bar{u} \in \mathcal{U},\ |u| \leq |\bar{u}| \Rightarrow u \in \mathcal{U}$$

where, as usual, $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^m$. This stronger property means that there exists a solid linear space $\mathcal{U}_0$ of real-valued random variables such that

$$\mathcal{U} = \{u \in L^0 \mid |u| \in \mathcal{U}_0\} \tag{4.3}$$

or, equivalently, that

$$\mathcal{U} = \mathcal{U}_0^m.$$

A benefit of our general definition of solidity is that it does not require all components of u to belong to the same space.

Proof. The stronger solidity property means that

$$\mathcal{U} = \{u \in L^0 \mid \exists \bar{u} \in \mathcal{U} :\ |u| \leq |\bar{u}|\ a.s.\}$$

which means that (4.3) holds with

$$\mathcal{U}_0 := \{\xi \in L^0 \mid \exists u \in \mathcal{U} :\ |\xi| \leq |u|\ a.s.\}.$$

Linearity and solidity together with (4.3) imply $\mathcal{U} = \mathcal{U}_0^m$. Assume now that $\mathcal{U} = \mathcal{U}_0^m$ for a linear solid $\mathcal{U}_0$. Let $u \in L^0$ and $\bar{u} \in \mathcal{U}$ with $|u| \leq |\bar{u}|$. There is a constant $c > 0$ such that

$$c \sum_{i=1}^m |u^i| \leq |u| \leq |\bar{u}| \leq \sum_{i=1}^m |\bar{u}^i|,$$

where $\sum_i |\bar{u}^i| \in \mathcal{U}_0$, by linearity. Thus, for each i, $|u^i| \leq \sum_i |\bar{u}^i|/c$ so $u^i \in \mathcal{U}_0$, by solidity. ☐

Most spaces of random variables encountered in applications are solid and decomposable. Examples include the classical Lebesgue, Orlicz and Lorentz spaces as well as spaces of finite moments. Cartesian products of solid and decomposable spaces are solid and decomposable. If $\mathcal{U}_i$ is in separating duality with $\mathcal{Y}_i$, then $\mathcal{U}_1 \times \cdots \times \mathcal{U}_m$ is in separating duality with $\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$. Spaces that do not satisfy our assumptions include the spaces of continuous functions or various Sobolev spaces of functions on $\mathbb{R}^n$ as they are neither decomposable nor solid. The space L^0 of all random variables is decomposable and solid, but if $(\Omega, \mathcal{F}, P)$ is atomless, it cannot be paired with a nontrivial space of random variables. Indeed, if $y \in L^0$ is nonzero, then, by **??**, $u \to E[u \cdot y]$ is improper on L^0.

In applications, one is often given a decomposable space $\mathcal{U} \subseteq L^1$ of random variables and then needs to find an appropriate dual space $\mathcal{Y}$. In general, there are several possibilities. The smallest decomposable space that is in separating duality with $\mathcal{U}$ with respect to the bilinear form $\langle u, y \rangle = E[u \cdot y]$ is the space L^∞ of essentially bounded functions. The largest one is the *Köthe dual*

$$\mathcal{U}' := \{y \in L^0 \mid u \cdot y \in L^1 \quad \forall u \in \mathcal{U}\}.$$

Köthe duals have simple characterizations for many familiar spaces of random variables. For example, $(L^p)' = L^q$ for any $p \in [1, \infty]$ and the usual conjugate exponent q of p. The following is easily verified.

Remark 4.2. *Given a linear space $\mathcal{U}$ of random variables, we have*

1. *If $L^\infty \subseteq \mathcal{U}$, then $\mathcal{U}' \subseteq L^1$;*
2. *If $\mathcal{U} \subseteq L^1$, then $L^\infty \subseteq \mathcal{U}'$;*
3. *If $u1_A \in \mathcal{U}$ for all $u \in \mathcal{U}$ and $A \in \mathcal{F}$, then $y1_A \in \mathcal{U}'$ for all $y \in \mathcal{U}'$ and $A \in \mathcal{F}$;*
4. *If $\mathcal{U}$ is solid, then $\mathcal{U}'$ is solid.*

In particular, if $\mathcal{U} \subseteq L^1$ is solid and decomposable, then $\mathcal{U}' \subseteq L^1$ is solid and decomposable.

A solid space containing all constant functions is decomposable. The following shows that the converse does not hold in general.

Example 4.3. *Let $u \geq 1$ be an unbounded real-valued random variable and let $\mathcal{U}$ be the sum of L^∞ and the linear span of the set $\{u1_A \mid A \in \mathcal{F}\}$. Then $\mathcal{U}$ is decomposable, by construction, but not solid, since it does not contain $\sqrt{u}$. Indeed, assume that*

$$\sqrt{u} = \bar{u} + \sum_{\nu=1}^{N} \alpha^\nu u 1_{A^\nu}$$

for some $\bar{u} \in L^\infty$, a finite partition $(A^\nu)_{\nu=1}^N$ and $\alpha^\nu \in \mathbb{R}$. Since $\sqrt{u}$ is unbounded, $1_{A^\nu} u$ has to be unbounded for some ν. We have

$$\sqrt{u} - \alpha^\nu u = \bar{u}$$

on A^ν, which is impossible if the left side is unbounded.

Given a topology on $\mathcal{U}$, the corresponding *topological dual* of $\mathcal{U}$ is the linear space of all continuous linear functionals on $\mathcal{U}$. A topology is *compatible* with the bilinear form on $\mathcal{U} \times \mathcal{Y}$ if every continuous linear functional on $\mathcal{U}$ can be expressed in the form

$$u \mapsto \langle u, y \rangle$$

for some $y \in \mathcal{Y}$. Such topologies can be characterized in terms of the "weak" and "Mackey" topologies associated with the bilinear form. The *weak topology* $\sigma(\mathcal{U}, \mathcal{Y})$ on $\mathcal{U}$ is the topology generated by the linear functionals $u \mapsto \langle u, y \rangle$ where $y \in \mathcal{Y}$. Similarly for $\mathcal{Y}$. The *Mackey topology* $\tau(\mathcal{U}, \mathcal{Y})$ is the topology generated by the sublinear functionals

$$\sigma_D(u) := \sup_{y \in D} \langle u, y \rangle,$$

where $D \subset \mathcal{Y}$ is $\sigma(\mathcal{Y},\mathcal{U})$-compact. Similarly for $\mathcal{Y}$. Given a topology on $\mathcal{U}$, the corresponding topological dual can be identified with $\mathcal{Y}$ if and only if the topology is between $\sigma(\mathcal{U}, \mathcal{Y})$ and $\tau(\mathcal{U}, \mathcal{Y})$. If $\mathcal{U}$ is a Fréchet (e.g. Banach) space under a given topology s and if $\mathcal{Y}$ the topological dual of $\mathcal{U}$, then $\tau(\mathcal{U}, \mathcal{Y})$ coincides with s. In particular if $\mathcal{U} = L^p$ and $\mathcal{Y} = L^q$ with $p \in [1, \infty)$ and q the conjugate exponent of p, then the Mackey topology $\tau(\mathcal{U}, \mathcal{Y})$ is just the usual L^p norm topology. However, $\tau(L^\infty, L^1)$ is, in general, strictly smaller than the norm topology of L^∞.

The following relates the weak and Mackey topologies on $\mathcal{U}$ to those on L^1 and L^∞ as well as to the metric topology of L^0.

Lemma 4.4. *We have $L^\infty \subseteq \mathcal{U} \subseteq L^1$ and $L^\infty \subseteq \mathcal{Y} \subseteq L^1$ and*

$$\begin{aligned} \sigma(L^1, L^\infty)|_{\mathcal{U}} \subseteq \sigma(\mathcal{U}, \mathcal{Y}), \quad \sigma(\mathcal{U}, \mathcal{Y})|_{L^\infty} \subseteq \sigma(L^\infty, L^1), \\ \tau(L^1, L^\infty)|_{\mathcal{U}} \subseteq \tau(\mathcal{U}, \mathcal{Y}), \quad \tau(\mathcal{U}, \mathcal{Y})|_{L^\infty} \subseteq \tau(L^\infty, L^1). \end{aligned}$$

The L^0-topology on $\mathcal{U}$ is weaker than $\tau(\mathcal{U}, \mathcal{Y})$.

Proof. Since $\mathcal{U}$ and $\mathcal{Y}$ are decomposable, $L^\infty \subseteq \mathcal{U}$ and $L^\infty \subseteq \mathcal{Y}$. Let $u \in \mathcal{U}$ and define $y \in L^\infty$ by $y^i = \operatorname{sign}(u^i)$. We have $\|u\|_{L^1} = E[u \cdot y] \in \mathbb{R}$. Thus, $\mathcal{U} \subseteq L^1$, and, by symmetry, $\mathcal{Y} \subseteq L^1$. The inclusions $L^\infty \subseteq \mathcal{U} \subseteq L^1$ and $L^\infty \subseteq \mathcal{Y} \subseteq L^1$ give the relations for the σ-topologies. Since, by symmetry, analogous relations are valid for the σ-topologies on $\mathcal{Y}$, $\sigma(L^\infty, L^1)$-compact subsets of L^∞ are $\sigma(\mathcal{Y},\mathcal{U})$-compact. Since $\tau(\mathcal{U}, \mathcal{Y})$ is generated by the support functions of $\sigma(\mathcal{Y},\mathcal{U})$-compact sets, we get $\tau(L^1, L^\infty)|_{\mathcal{U}} \subseteq \tau(\mathcal{U}, \mathcal{Y})$. The remaining inclusion is verified similarly. As noted above, $\tau(L^1, L^\infty)$-topology is the L^1-norm topology on L^1. Since the L^0-topology on L^1 is weaker than the norm topology, the last claim follows from $\tau(L^1, L^\infty)|_{\mathcal{U}} \subseteq \tau(\mathcal{U}, \mathcal{Y})$. □

Let $\mathcal{X}$ and $\mathcal{V}$ be decomposable solid spaces of $\mathbb{R}^n$-valued random variables in separating duality under the bilinear form

$$\langle x, v \rangle := E[x \cdot v].$$

A linear mapping $\mathcal{A} : \mathcal{X} \to \mathcal{U}$ is *weakly continuous* if it is continuous with respect to the weak topologies on $\mathcal{X}$ and $\mathcal{U}$. This means that $x \mapsto \langle \mathcal{A}x, y\rangle$ is $\sigma(\mathcal{X}, \mathcal{V})$-continuous for all $y \in \mathcal{Y}$, or equivalently, there exists a linear mapping $\mathcal{A}^* : \mathcal{Y} \to \mathcal{V}$ such that

$$\langle \mathcal{A}x, y\rangle = \langle x, \mathcal{A}^* y\rangle \quad \forall x \in \mathcal{X},\ y \in \mathcal{Y}.$$

The mapping $\mathcal{A}^*$ is known as the *adjoint* of $\mathcal{A}$.

We use the fact that the scenariowise Moore-Penrose inverse $A^\dagger$ of a random matrix A is measurable.

Lemma 4.5. *Let $A \in L^0(\mathbb{R}^{m\times n})$ be a random matrix such that $Ax \in \mathcal{U}$ for all $x \in \mathcal{X}$. The linear mapping $\mathcal{A} : \mathcal{X} \to \mathcal{U}$ defined pointwise by*

$$\mathcal{A}x = Ax \quad a.s.$$

is weakly continuous if and only if $A^* y \in \mathcal{V}$ for all $y \in \mathcal{Y}$, and in this case its adjoint is given pointwise by

$$\mathcal{A}^* y = A^* y \quad a.s.,$$

the weak closure of $\operatorname{rge} \mathcal{A}$ *is* $\mathcal{U}(\operatorname{rge} A)$ *and if $A^\dagger u \in \mathcal{X}$ for all $u \in \mathcal{U}$, then* $\operatorname{rge} \mathcal{A}$ *is weakly closed in* $\mathcal{U}$*. If $\mathcal{V}$ is the Köthe dual of $\mathcal{X}$, then $A^* y \in \mathcal{V}$ for all $y \in \mathcal{Y}$.*

Proof. For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\langle \mathcal{A}x, y\rangle = E[(Ax) \cdot y] = E[x \cdot A^* y],$$

which proves the equivalence and the adjoint formula. When $\mathcal{V}$ is the Köthe dual of $\mathcal{X}$, the above equation implies that $x \cdot A^* y \in L^1$ for all $x \in \mathcal{X}$, so $A^* y \in \mathcal{V}$.

It is clear that $\operatorname{rge} \mathcal{A} \subseteq \mathcal{U}(\operatorname{rge} A)$. By **??**, the set $\mathcal{U}(\operatorname{rge} A)$ is L^0-closed so, by 4.4, it is also weakly closed. It follows that $\operatorname{cl} \operatorname{rge} \mathcal{A} \subseteq \mathcal{U}(\operatorname{rge} A)$. Given $u \in \mathcal{U}(\operatorname{rge} A)$, there exists, by 3.15, an $x \in L^0$ with $u = Ax$. Defining $x^\nu := x 1_{\{|x|\le\nu\}}$, we have $x^\nu \in \mathcal{X}$ and $Ax^\nu = u 1_{\{|x|\le\nu\}}$, so

$$E[Ax^\nu \cdot y] \to E[u \cdot y] \quad \forall y \in \mathcal{Y},$$

by dominated convergence. Thus, $Ax^\nu \to u$ weakly, so $\operatorname{cl} \operatorname{rge} \mathcal{A} = \mathcal{U}(\operatorname{rge} A)$.

If $A^\dagger u \in \mathcal{X}$ for all $u \in \mathcal{U}$, then any $u \in \operatorname{rge} \mathcal{A}$ can be expressed as $u = \mathcal{A}(\mathcal{A}^\dagger u)$, where $\mathcal{A}^\dagger : \mathcal{U} \to \mathcal{X}$ is defined pointwise by $(\mathcal{A}^\dagger u)(\omega) := A^\dagger(\omega) u(\omega)$. By the first claim, $\mathcal{A}$ and $\mathcal{A}^\dagger$ are both weakly continuous. The set $\operatorname{rge} \mathcal{A}$ is thus closed since it is the kernel of the continuous mapping $u \mapsto \mathcal{A}\mathcal{A}^\dagger u - u$. □

The following characterizes the adjoint of the conditional expectation operator with respect to a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$; see Section **??**.

Lemma 4.6. *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra such that $E^{\mathcal{G}}u \in \mathcal{U}$ for all $u \in \mathcal{U}$. The mapping $E^{\mathcal{G}} : \mathcal{U} \to \mathcal{U}$ is weakly continuous if and only if $E^{\mathcal{G}}y \in \mathcal{Y}$ for all $y \in \mathcal{Y}$ and, in this case, its adjoint is given by*

$$(E^{\mathcal{G}})^* y = E^{\mathcal{G}} y \quad a.s.$$

If $\mathcal{Y}$ is the Köthe dual of $\mathcal{U}$, then $E^{\mathcal{G}}y \in \mathcal{Y}$ for all $y \in \mathcal{Y}$.

Proof. If u^i, y^i, $(E^{\mathcal{G}}u)^i y^i$ and $u^i(E^{\mathcal{G}}y)^i$ are integrable, **??** gives

$$E[E^{\mathcal{G}}u \cdot y] = E[(E^{\mathcal{G}}u) \cdot E^{\mathcal{G}}y] = E[u \cdot E^{\mathcal{G}}y]. \tag{4.4}$$

Thus, if $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$ and $E^{\mathcal{G}}\mathcal{Y} \subset \mathcal{Y}$, then, by (4.2), the function $u \mapsto E^{\mathcal{G}}u$ is weakly continuous. On the other hand, if $E^{\mathcal{G}} : \mathcal{U} \to \mathcal{U}$ is weakly continuous, then $u \mapsto E[E^{\mathcal{G}}u \cdot y]$ is $\sigma(\mathcal{U}, \mathcal{Y})$-continuous for $y \in \mathcal{Y}$. Thus, there exists a $y' \in \mathcal{Y}$ such that $E[E^{\mathcal{G}}u \cdot y] = E[u \cdot y']$ for all $u \in \mathcal{U}$. Since $y \in L^1$, (4.4) gives

$$E[E^{\mathcal{G}}u \cdot y] = E[u \cdot E^{\mathcal{G}}y] \quad \forall u \in L^\infty.$$

Thus, $y' = E^{\mathcal{G}}y$ almost surely so $E^{\mathcal{G}}\mathcal{Y} \subset \mathcal{Y}$.

Assume now that $\mathcal{Y}$ is the Köthe dual of $\mathcal{U}$ and let $y \in \mathcal{Y}$. It suffices to show that $E^{\mathcal{G}}y \in \mathcal{Y}$. By (4.2), it suffices to treat the case where at most one component y^i of y is nonzero. Without loss of generality, we may assume that y^i is nonnegative so that $E^{\mathcal{G}}y^i$ is nonnegative as well. If $\mathcal{Y}$ is the Köthe dual, it suffices to show that $E[u^i(E^{\mathcal{G}}y^i)] < \infty$ for every nonnegative $u \in \mathcal{U}$. By **??**, $E[u^i(E^{\mathcal{G}}y^i)] = E[E^{\mathcal{G}}(u^i)y^i]$, where the right side is finite, since $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$. □

Many familiar spaces of random variables satisfy the condition $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$ in 4.6 for every σ-algebra $\mathcal{G} \subseteq \mathcal{F}$; see **??**. The condition may fail, e.g., in Musielak-Orlicz spaces with random Young functions.

When the matrix A in 4.5 is $\mathcal{G}$-measurable, then the corresponding mapping $\mathcal{A} : \mathcal{X} \to \mathcal{U}$ commutes with the $\mathcal{G}$-conditional expectation.

Lemma 4.7. *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra such that $E^{\mathcal{G}}u \in \mathcal{U}$ for all $u \in \mathcal{U}$ and let $A \in L^0(\mathbb{R}^{m\times n})$ be a $\mathcal{G}$-measurable random matrix such that $Ax \in \mathcal{U}$ for all $x \in \mathcal{X}$. Then*

$$E^{\mathcal{G}}[Ax] = AE^{\mathcal{G}}x$$

for all $x \in \mathcal{X}$.

Proof. Let $x \in \mathcal{X}$. Changing x by setting all but its jth component to zero, we still have $x \in \mathcal{X}$, by solidity of $\mathcal{X}$. Thus, $u := (A_{ij}x^j)_{i=1}^m \in \mathcal{U}$ since $Ax \in \mathcal{U}$ for all $x \in \mathcal{X}$. Changing u by setting all but its ith component to zero, we still have $u \in \mathcal{U}$, by solidity of $\mathcal{U}$. Thus, by 4.4, $A_{i,j}x^j$ is integrable for every i and j. By **??**.2, $E^{\mathcal{G}}[A_{ij}x^j] = A_{ij}E^{\mathcal{G}}x^j$. Applying **??**.1 to each component of Ax then gives $E^{\mathcal{G}}[Ax] = AE^{\mathcal{G}}x$. □

4.2 Conjugates of integral functionals

This section studies convex integral functionals on paired decomposable spaces $\mathcal{U}$ and $\mathcal{Y}$ of random variables. More precisely, we fix a normal integrand h and study the integral functionals $Eh : \mathcal{U} \to \overline{\mathbb{R}}$ and $Eh^* : \mathcal{Y} \to \overline{\mathbb{R}}$ defined by

$$Eh(u) := \int_\Omega h(u(\omega), \omega) dP(\omega)$$

and

$$Eh^*(y) := \int_\Omega h^*(y(\omega), \omega) dP(\omega),$$

where

$$h^*(v, \omega) := \sup_{x \in \mathbb{R}^m} \{x \cdot v - h(x, \omega)\}.$$

By 3.26, h^* is a convex normal integrand. The next result characterizes the conjugate and the subdifferential of Eh with respect to the pairing of $\mathcal{U}$ with $\mathcal{Y}$; see Section **??**. Recall that the conjugate of f is defined for each $y \in \mathcal{Y}$ by

$$(Eh)^*(y) := \sup_{u \in \mathcal{U}} \{\langle u, y \rangle - Eh(u)\}$$

while the subdifferential $\partial Eh(u)$ of Eh at a $u \in \mathcal{U}$ is the closed convex set

$$\partial Eh(u) := \{y \in \mathcal{Y} \mid Eh(u') \geq Eh(u) + \langle u' - u, y \rangle \quad \forall u' \in \mathcal{U}\}.$$

Given $u \in \mathcal{U}$, the mapping

$$\omega \mapsto \partial h(u)(\omega) := \partial h(u(\omega), \omega)$$

is measurable, by 3.27.

Theorem 4.8. *If h is a convex normal integrand with* $\operatorname{dom} Eh \neq \emptyset$*, then*

$$(Eh)^* = Eh^*$$

and

$$\partial Eh(u) = \{y \in \mathcal{Y} \mid y \in \partial h(u) \ a.s.\}$$

for any $u \in \mathcal{U}$ such that $Eh(u)$ is finite.

Proof. By 3.33,

$$\langle u, y \rangle - Eh(u) = E[u \cdot y - h(u)]$$

for every $u \in \mathcal{U}$ and $y \in \mathcal{Y}$. The first claim thus follows by applying 3.29 to the normal integrand $h_y(u, \omega) := h(u, \omega) - u \cdot y(\omega)$. As to the second, we have, by definition, $y \in \partial Eh(u)$ if and only if $Eh(u) + (Eh)^*(y) = \langle u, y \rangle$. By the first claim, this is equivalent to

$$Eh(u) + Eh^*(y) = \langle u, y \rangle.$$

By Fenchel's inequality, $h(u)+h^*(y) \geq u\cdot y$ almost surely so, by 3.33, $y \in \partial Eh(u)$ if and only if

$$E[h(u) + h^*(y) - u \cdot y] = 0.$$

By Fenchel's inequality again, this holds if and only if

$$h(u) + h^*(y) = u \cdot y$$

almost surely. This means that $y \in \partial h(u)$ almost surely. □

Recall that a convex function g in a locally convex vector space is lsc with respect to the weak topology if it is lsc merely with respect to the Mackey topology; see **??**. The converse is immediate. From now on, we will simply say that a convex function g on a locally convex vector space is *lsc* if it is lower semicontinuous with respect to the Mackey topology. Accordingly, we say that a convex set is *closed* if it is closed with respect to the Mackey topology. The *closure* of a function g is defined by

$$\operatorname{cl} g = \begin{cases} \operatorname{lsc} g & \text{if } \operatorname{lsc} g(u) > -\infty \text{ for all } u \in U, \\ -\infty & \text{otherwise;} \end{cases}$$

see Section **??**. By 2.23, $\operatorname{cl} g = g^{**}$. The function g is said to be *closed* a point u if $g(u) = (\operatorname{cl} g)(u)$. A function which is closed at every point is said to be *closed*. Clearly, a convex set is closed if and only if its indicator closed.

Corollary 4.9. *Let h be a convex normal integrand. The following are equivalent:*

1. $\operatorname{dom} Eh \neq \emptyset$ *and* $\operatorname{dom} Eh^* \neq \emptyset$*;*

2. *Eh is closed and proper;*

3. *Eh^* is closed and proper;*

4. $\operatorname{dom} Eh \neq \emptyset$ *and there exist $y \in \mathcal{Y}$ and $\alpha \in L^1$ such that*

$$h(u, \omega) \geq u \cdot y(\omega) - \alpha(\omega) \quad \forall u \in \mathbb{R}^m;$$

5. $\operatorname{dom} Eh^* \neq \emptyset$ *and there exist $u \in \mathcal{U}$ and $\alpha \in L^1$ such that*

$$h^*(y, \omega) \geq u(\omega) \cdot y - \alpha(\omega) \quad \forall u \in \mathbb{R}^m$$

and imply that Eh and Eh^* are conjugates of each other and that $y \in \partial Eh(u)$ if and only if $y \in \partial h(u)$ almost surely.

Proof. We prove the equivalence of 1, 2 and 4. The equivalence with 3 and 5 then follow by symmetry. By 4.8, 1 implies that Eh and Eh^* are conjugates of each other so both are closed and proper and thus, 2 and 3 hold. Assuming 2, the biconjugate theorem gives the existence of $y \in \mathcal{Y}$ and $a \in \mathbb{R}$ such that

$$Eh(u) \geq \langle u, y \rangle - a \quad \forall u \in \mathcal{U}.$$

Thus, by 4.8
$$a \geq (Eh)^*(y) = Eh^*(y).$$
By Fenchel's inequality,
$$h(u,\omega) + h^*(y,\omega) \geq u \cdot y,$$
so 4 holds with $\alpha(\omega) = h^*(y(\omega),\omega)$. If 4 holds, $Eh^*(y) \leq E\alpha$, so 1 holds.

Assume 1. By 4.8, $y \in \partial Eh(u)$ implies $y \in \partial h(u)$ almost surely. On the other hand, if $u \in \mathcal{U}$ and $y \in \mathcal{Y}$ are such that $y \in \partial h(u)$ almost surely, then $h(u) + h^*(y) = u \cdot y$ almost surely. The properness of Eh and Eh^* implies that the negative parts of $h(u)$ and $h^*(y)$ are integrable so, by 3.33, $Eh(u)+Eh^*(y) = \langle u, y\rangle$, which means that $y \in \partial Eh(u)$. □

Applying 4.9 to the indicator function of a closed-valued measurable mapping, gives the following.

Corollary 4.10. *Given a closed convex-valued measurable mapping $S : \Omega \rightrightarrows \mathbb{R}^m$, the set*
$$\mathcal{U}(S) := \{u \in \mathcal{U} \mid u \in S \ a.s.\}$$
is closed and convex.

Proof. If $\mathcal{U}(S) = \emptyset$ the claim holds trivially. If $\mathcal{U}(S) \neq \emptyset$ it follows by applying 4.9 to $h(u,\omega) := \delta_S(u,\omega)$. Indeed, we have $0 \in \operatorname{dom} Eh^*$ so condition 1 of 4.9 holds. Thus, the function $Eh = \delta_{\mathcal{U}(S)}$ is closed. □

Given a convex set $C \subset \mathcal{U}$, its *core* is the set $\operatorname{core} C$ of points $u \in C$ such that the *positive hull*
$$\operatorname{pos}(C - u) := \bigcup_{\lambda > 0} \lambda(C - u)$$
is the whole space $\mathcal{U}$. Recall that the Köthe dual of $\mathcal{U}$ is the linear space
$$\mathcal{U}' := \{y \in L^0 \mid u \cdot y \in L^1 \quad \forall u \in \mathcal{U}\}.$$

Theorem 4.11. *If Eh is proper and $\mathcal{Y} = \mathcal{U}'$, then*
$$\partial Eh(u) = \{y \in L^0 \mid y \in \partial h(u) \ a.s.\} \neq \emptyset$$
for all $u \in \operatorname{core} \operatorname{dom} Eh$ and Eh is closed as soon as $\operatorname{core} \operatorname{dom} Eh \neq \emptyset$.

Proof. Let $u \in \operatorname{core} \operatorname{dom} Eh$. We have $u \in \operatorname{int} \operatorname{dom} h$ almost surely. Indeed, given a finite set $\{w_i\}_{i \in I} \subset \mathbb{R}^m$ whose convex hull is a neighborhood of the origin, there is an $\epsilon > 0$ such that $u + \epsilon w_i \in \operatorname{dom} Eh$ for all $i \in I$. Thus, $u + \epsilon w_i \in \operatorname{dom} h$ for all $i \in I$ almost surely so $u \in \operatorname{int} \operatorname{dom} h$ almost surely.

By **??** and **??**, $\partial h(u) \neq \emptyset$ almost surely. By the measurable selection theorem, there exists a $y \in L^0$ with $y \in \partial h(u)$ almost surely, i.e.

$$h(u(\omega) + u', \omega) \geq h(u(\omega), \omega) + u' \cdot y(\omega) \quad \forall u' \in \mathbb{R}^m.$$

Given any $u' \in \mathcal{U}$ and $\beta > 0$, this implies,

$$E[u' \cdot y] \leq \frac{Eh(u + \beta u') - Eh(u)}{\beta}.$$

Since $u \in \operatorname{core} \operatorname{dom} Eh$, there is a $\beta > 0$ such that the right side is finite. Thus, $y \in \mathcal{U}'$ so $y \in \mathcal{Y}$, by assumption. Together with the above inequalities, this proves the first claim. The condition $y \in \partial h(u)$ means that

$$h^*(y) = u \cdot y - h(u)$$

so we also get that both Eh and Eh^* have nonempty domains. The last claim thus follows from 4.9. □

The *relative core* of a set $C \subset \mathcal{U}$ is the core of C relative to its affine hull; see Section **??**. By **??**, $\operatorname{rcore} C$ is set of points $u \in C$ such that $\operatorname{pos}(C - u)$ is linear.

Remark 4.12. *Assume that Eh is proper,* $\mathcal{Y} = \mathcal{U}'$,

$$\mathcal{U}(\operatorname{aff} \operatorname{dom} h) \subseteq \operatorname{aff} \operatorname{dom} Eh,$$

and that $\mathcal{U}$ *satisfies the stronger solidity property in 4.1. Then*

$$\partial Eh(u) \neq \emptyset$$

for all $u \in \operatorname{rcore} \operatorname{dom} Eh$ *and* Eh *is closed as soon as* $\operatorname{rcore} \operatorname{dom} Eh \neq \emptyset$.

Proof. Let $u \in \operatorname{rcore} \operatorname{dom} Eh$ and let π be the scenariowise projection to $\operatorname{aff} \operatorname{dom} h - u$. We have $u \in \operatorname{rint} \operatorname{dom} h$ almost surely. Indeed, let $\{w_i\}_{i \in I} \subset \mathbb{R}^m$ be a finite set whose convex hull is a neighborhood of the origin in $\mathbb{R}^m$. Since $\mathcal{U}(\operatorname{aff} \operatorname{dom} h) \subset \operatorname{aff} \operatorname{dom} Eh$ by assumption, there is an $\epsilon > 0$ such that $u + \epsilon \pi w_i \in \operatorname{dom} Eh$ for all $i \in I$. It follows that $u + \epsilon \pi w_i \in \operatorname{dom} h$ for all $i \in I$ almost surely so $u \in \operatorname{rint} \operatorname{dom} h$ almost surely. By **??** and **??**, $\partial h(u) \neq \emptyset$ almost surely. By the measurable selection theorem, there exists a $y \in L^0$ with $y \in \partial h(u)$ almost surely, i.e.

$$h(u(\omega) + u', \omega) \geq h(u(\omega), \omega) + u' \cdot y(\omega) \quad \forall u' \in \mathbb{R}^m. \tag{4.5}$$

Clearly, $\pi y \in \partial h(u)$ as well.

Let $u' \in \mathcal{U}$. Since, by **??**, $|\pi u'| \leq |u'|$ almost surely, we have $\pi u' \in \mathcal{U}$ since $\mathcal{U}$ satisfies the stronger solidity property in 4.1. By assumption, $\mathcal{U}(\operatorname{aff} \operatorname{dom} h) \subset \operatorname{aff} \operatorname{dom} Eh$, so there exists $\lambda > 0$ such that $u + \lambda \pi u' \in \operatorname{dom} Eh$. Combining this with (4.5) gives

$$E[\lambda u' \cdot \pi y] = E[\lambda \pi u' \cdot y] \leq Eh(u + \lambda \pi u') - Eh(u) < \infty.$$

Since $u' \in \mathcal{U}$ was arbitrary, this implies that πy is in the Köthe dual of $\mathcal{U}$ and thus, by assumption, in $\mathcal{Y}$. This proves the first claim. The last claim follows from the conditions $\pi y \in \partial h(u)$ and $\pi y \in \mathcal{Y}$ just like in the proof of 4.11. □

Note that the last assumption in 4.12 is slightly stronger than mere solidity of $\mathcal{U}$ which means that $u \in \mathcal{U}$ for all $u \in L^0$ such that $|u^i| \le |\bar{u}^i|$ for some $\bar{u} \in \mathcal{U}$.

The following is a corollary of **??**.

Theorem 4.13 (Jensen's inequality)**.** *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra such that $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$ and $E^{\mathcal{G}}\mathcal{Y} \subset \mathcal{Y}$ and let h be a $\mathcal{G}$-measurable convex normal integrand such that $Eh^*(y) < \infty$ for some $y \in \mathcal{Y}$. Then*

$$Eh(E^{\mathcal{G}}u) \le Eh(u)$$

for every $u \in \mathcal{U}$.

Proof. By Fenchel's inequality,

$$Eh(u) \ge E[u \cdot y] - Eh^*(y)$$

for all $u \in \mathcal{U}$, so the claim follows from **??**. □

The following extends **??** from L^1 to more general $\mathcal{U}$.

Theorem 4.14. *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra such that $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$ and $E^{\mathcal{G}}\mathcal{Y} \subset \mathcal{Y}$. Given a closed convex-valued random set S with $\mathcal{U}(S) \neq \emptyset$, the following are equivalent:*

1. *S is $\mathcal{G}$-measurable;*
2. *$E^{\mathcal{G}}u \in \mathcal{U}(S)$ for every $u \in \mathcal{U}(S)$;*
3. *$E^{\mathcal{G}}S \subseteq S$ almost surely.*

Proof. By **??**, it suffices to show that condition 2 here implies that of **??**. Condition 2 means that $E\delta_S(E^{\mathcal{G}}u) \le E\delta_S(u)$ for all $u \in \mathcal{U}$. By 4.9, **??** and 4.6, this means that $E\sigma_S(E^{\mathcal{G}}y) \le E\sigma_S(y)$ for all $y \in \mathcal{Y}$. In particular,

$$E\sigma_S(E^{\mathcal{G}}y) \le E\sigma_S(y) \quad \forall y \in L^\infty.$$

Applying the same argument in the pairing of L^1 with L^∞ now gives

$$E\delta_S(E^{\mathcal{G}}u) \le E\delta_S(u) \quad \forall u \in L^1,$$

which is condition 2 in **??**. □

The following is a corollary of **??**.

Theorem 4.15. *Let h be a convex normal integrand such that $Eh : \mathcal{U} \to \overline{\mathbb{R}}$ is proper and closed. Then*

$$(Eh)^\infty = Eh^\infty,$$
$$\sigma_{\operatorname{dom} Eh} = E\sigma_{\operatorname{dom} h}$$

and

$$\operatorname{cl} \operatorname{dom} Eh = \mathcal{U}(\operatorname{cl} \operatorname{dom} h).$$

Proof. By 4.9, there exists $y \in \mathcal{Y}$ such that $h^*(y)$ is integrable. By Fenchel's inequality,

$$\bar{h}(u, \omega) := h(u, \omega) - u \cdot \bar{y}(\omega) \geq -h^*(y(\omega), \omega),$$

so, by 3.31, $E\bar{h}$ is proper and lsc on L^0. By **??**,

$$(E\bar{h})^\infty = E\bar{h}^\infty.$$

Since $\bar{h}^\infty(u, \omega) = h^\infty(u, \omega) - u \cdot y(\omega)$ and $(E\bar{h})^\infty(u) = (Eh)^\infty - E[u \cdot y]$ on $\mathcal{U}$, we get $(Eh)^\infty = Eh^\infty$ on $\mathcal{U}$. The second expression follows from the first one and **??**. Applying 4.9 and 2.23 to the second expression, we get $\operatorname{cl} \delta_{\operatorname{dom} Eh} = E\delta_{\operatorname{cl} \operatorname{dom} h}$, which is the last expression. □

5 Duality for integrable strategies

We now return to the problem

$$\text{minimize } \; Eh(x) := \int h(x(\omega), \omega) dP(\omega) \; \text{ over } x \in \mathcal{X} \cap \mathcal{N} \qquad (P_{\mathcal{X}})$$

from the introduction of this chapter. Again, we assume that $h(x, \omega) = f(x, \bar{u}(\omega), \omega)$ for a convex normal integrand f and random vector $\bar{u} \in \mathcal{U}$. By Theorem 3.9.4, such an h is a normal integrand. As observed in the introduction, $(P_{\mathcal{X}})$ fits the duality framework of Section 2.9 with the Rockafellian $F : \mathcal{X} \times \mathcal{X} \times \mathcal{U} \to \overline{\mathbb{R}}$ defined by

$$F(x, z, u) := Ef(x, u) + \delta_{\mathcal{N}}(x - z)$$

and the dualizing parameter $(z, u) \in \mathcal{X} \times \mathcal{U}$. Clearly,

$$F(x, z, u) = Ef(x, u) + \delta_{\mathcal{X}_a}(x - z) \quad \forall (x, z, u) \in \mathcal{X} \times \mathcal{X} \times \mathcal{U},$$

where

$$\mathcal{X}_a := \mathcal{X} \cap \mathcal{N}$$

is the linear space of the adapted strategies in $\mathcal{X}$.

In order to apply the results of Section 4, we assume that $\mathcal{X}$ and $\mathcal{U}$ are solid decomposable spaces in separating duality with solid decomposable spaces $\mathcal{V} \subseteq L^0(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ and $\mathcal{Y} \subseteq L^0(\Omega, \mathcal{F}, P; \mathbb{R}^m)$, respectively, under the bilinear forms

$$\langle x, v \rangle := E[x \cdot v] \quad \text{and} \quad \langle u, y \rangle := E[u \cdot y].$$

Solidity implies that

$$\mathcal{X} = \mathcal{X}_0 \times \cdots \times \mathcal{X}_T \quad \text{and} \quad \mathcal{V} = \mathcal{V}_0 \times \cdots \times \mathcal{V}_T,$$

where $\mathcal{X}_t$ and $\mathcal{V}_t$ are solid decomposable spaces of $\mathbb{R}^{n_t}$-valued random variables in separating duality under the bilinear form $(x_t, v_t) \mapsto E[x_t \cdot v_t]$. It follows that

$$\langle x, v\rangle = \sum_{t=0}^{T} E[x_t \cdot v_t] \quad \forall x \in \mathcal{X},\ v \in \mathcal{V}$$

and

$$\mathcal{X}_a = \mathcal{X}_0(\mathcal{F}_0) \times \cdots \times \mathcal{X}_T(\mathcal{F}_T).$$

We will denote the orthogonal complement of $\mathcal{X}_a$ by

$$\mathcal{X}_a^\perp := \{v \in \mathcal{V} \,|\, \langle x, v\rangle = 0 \quad \forall x \in \mathcal{X}_a\}.$$

The following generalizes **??** which characterizes the set

$$\mathcal{N}^\perp := \{v \in L^1 \mid \langle x, v\rangle = 0 \quad \forall x \in \mathcal{N} \cap L^\infty\}.$$

Lemma 5.1. *The set $\mathcal{X}_a$ is closed and*

$$\mathcal{X}_a^\perp = \mathcal{N}^\perp \cap \mathcal{V} = \{v \in \mathcal{V} \,|\, E_t v_t = 0 \quad t = 0, \dots, T\}.$$

Proof. By **??**, $\mathcal{N}$ is closed in L^0 so the first claim follows from 4.4. Since

$$\mathcal{X}_a = \mathcal{X}_0(\mathcal{F}_0) \times \cdots \times \mathcal{X}_T(\mathcal{F}_T),$$

we have $v \in \mathcal{X}_a^\perp$ if and only if $E[x_t \cdot v_t] = 0$ for every $x_t \in \mathcal{X}_t(\mathcal{F}_t)$. Here, $E[x_t \cdot v_t] = E[x_t \cdot (E_t v_t)]$, by **??**. The claim now follows from the fact that $\mathcal{X}_t(\mathcal{F}_t)$ separates points in $\mathcal{V}_t(\mathcal{F}_t)$. Indeed, since the spaces are decomposable, we have $L^\infty(\mathcal{F}_t) \subseteq \mathcal{X}_t(\mathcal{F}_t)$ and $\mathcal{V}_t(\mathcal{F}_t) \subseteq L^1(\mathcal{F}_t)$. □

According to the general duality framework of Section 2.9, the *dual problem* associated with the above specifications is the concave maximization problem

$$\text{maximize} \quad \langle \bar{u}, y\rangle - F^*(0, p, y) \quad \text{over} (p, y) \in \mathcal{V} \times \mathcal{Y}, \tag{D}$$

where $F^* : \mathcal{V} \times \mathcal{V} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is the conjugate of F, i.e.

$$F^*(v, p, y) := \sup_{x,z,u} \{\langle x, v\rangle + \langle z, p\rangle + \langle u, y\rangle - F(x, z, u)\}.$$

An explicit expression for F^* will be given in 5.5 below. By Fenchel's inequality,

$$F(x, 0, u) \geq \langle u, y\rangle - F^*(0, p, y) \quad \forall x \in \mathcal{X},\ u \in \mathcal{U},\ p \in \mathcal{V},\ y \in \mathcal{Y},$$

so

$$\inf (P_{\mathcal{X}}) \geq \sup (D),$$

where $\inf(P_{\mathcal{X}})$ and $\sup(D)$ denote the optimum values of $(P_{\mathcal{X}})$ and (D), respectively. A *duality gap* is said to exist if the inequality is strict. Conversely, we say that there is no duality gap if $\inf(P_{\mathcal{X}}) = \sup(D)$.

The *Lagrangian* associated with the function F is the convex-concave function L on $\mathcal{X} \times \mathcal{V} \times \mathcal{Y}$ given by

$$L(x,p,y) := \inf_{(z,u)\in\mathcal{X}\times\mathcal{U}} \{F(x,z,u) - \langle z,p\rangle - \langle u,y\rangle\}.$$

The associated minimax problem is to find a saddle value and/or a saddle point of the concave-convex function

$$L_{\bar{u}}(x,p,y) := L(x,p,y) + \langle \bar{u}, y\rangle$$

when minimizing over x and maximizing over (p,y). If

$$\inf_x \sup_{p,y} L_{\bar{u}}(x,p,y) = \sup_{p,y} \inf_x L_{\bar{u}}(x,p,y),$$

the common value is called the *saddle value* of $L_{\bar{u}}$. A point (x,p,y) is a *saddle point* of $L_{\bar{u}}$ if

$$L_{\bar{u}}(x,p',y') \le L_{\bar{u}}(x,p,y) \le L_{\bar{u}}(x',p,y) \quad \forall x' \in \mathcal{X},\ p' \in \mathcal{V},\ y' \in \mathcal{Y}.$$

Clearly, the existence of a saddle point implies the existence of a saddle value. By definition, the conjugate of F can be expressed in term of the Lagrangian as

$$F^*(v,p,y) = \sup_{x\in\mathcal{X}} \{\langle x,v\rangle - L(x,p,y)\}.$$

It follows that the dual problem coincides with the maximization half of the minimax problem. Similarly, if $F(x,z,u)$ is closed in (z,u), then by 2.23, the primal problem coincides with the minimization half of the minimax problem.

The next three theorems are direct consequences of 2.31 and 2.32 and 2.33 in the appendix. They all involve the assumption that the integral functional Ef be closed in u. This means that $Ef(x,\cdot)$ is closed in $\mathcal{U}$ for each $x \in \mathcal{X}$. Combined with 5.1, this implies that the function F is closed in (z,u). Recall that, by the biconjugate theorem 2.23, a convex function is closed if and only if it coincides with its biconjugate.

The *optimum value function* $\varphi : \mathcal{X} \times \mathcal{U} \to \overline{\mathbb{R}}$ associated with F will be denoted by

$$\varphi(z,u) := \inf_{x\in\mathcal{X}} F(x,z,u).$$

By definition of F,

$$\varphi(z,u) = \inf_{x\in\mathcal{X}} \{Ef(x,u) \mid x - z \in \mathcal{N}\} = \inf_{x\in\mathcal{X}} \{Ef(x,u) \mid x - z \in \mathcal{X}_a\}.$$

Clearly, $\varphi(0,\bar{u}) = \inf(P_{\mathcal{X}})$. Note that we deviate slightly from the notation of Section 2.9 where the primal optimum value function has a subindex v. We omit the subindex here since we set $v = 0$ throughout. Clearly, $\varphi^*(p, y) = F^*(0, p, y)$ so the dual problem can be written also as

$$\text{maximize} \quad \langle \bar{u}, y\rangle - \varphi^*(p, y) \quad \text{over}(p, y) \in \mathcal{V} \times \mathcal{Y}.$$

The following is a direct consequence of 2.31.

Theorem 5.2. *The implications $1 \Leftrightarrow 2 \Rightarrow 3$ hold among the following conditions:*

1. $\inf (P_{\mathcal{X}}) = \sup (D)$;
2. φ *is closed at* $(0, \bar{u})$;
3. *The function* $L_{\bar{u}}$ *has a saddle value.*

If $Ef(x, u)$ *is closed in* u, *then* $1 \Leftrightarrow 2 \Leftrightarrow 3$.

The integral functional $Ef(x, u)$ is closed in u in particular if it is jointly closed in (x, u). By 4.8, this happens if $\operatorname{dom} Ef^* \cap (\mathcal{V} \times \mathcal{Y}) \neq \emptyset$.

The following restatement of 2.32 characterizes situations where there is no duality gap and, furthermore, the dual admits solutions. Recall that the subdifferential $\partial\varphi(z, u)$ of φ at a point $(z, u) \in \mathcal{X} \times \mathcal{U}$ is the closed convex set of points $(v, y) \in \mathcal{V} \times \mathcal{Y}$ such that

$$\varphi(z', u') \geq \varphi(z, u) + \langle z' - z, v\rangle + \langle u' - u, y\rangle \quad \forall (z', u') \in \mathcal{X} \times \mathcal{U}.$$

Theorem 5.3. *If* $\varphi(0, u) < \infty$, *then the implications* $1 \Leftrightarrow 2 \Rightarrow 3$ *hold among the following conditions:*

1. (p, y) *solves* (D) *and* $\inf (P_{\mathcal{X}}) = \sup (D)$;
2. *either* $(p, y) \in \partial\varphi(0, \bar{u})$ *or* $\varphi(0, \bar{u}) = -\infty$;
3. $\inf_x \sup_{p,y} L_{\bar{u}}(x, p, y) = \inf_x L_{\bar{u}}(x, p, y)$.

If, in addition, $Ef(x, u)$ *is closed in* u, *then* $1 \Leftrightarrow 2 \Leftrightarrow 3$.

The following restatement of 2.33 characterizes the situations where both primal and dual solutions exist and there is no duality gap.

Theorem 5.4. *The implications $1 \Leftrightarrow 2 \Rightarrow 3$ hold among the following conditions:*

1. x *solves* $(P_{\mathcal{X}})$, (p, y) *solves* (D) *and* $\inf (P_{\mathcal{X}}) = \sup (D) \in \mathbb{R}$;
2. $(0, p, y) \in \partial F(x, 0, \bar{u})$;
3. $0 \in \partial_x L(x, p, y)$ *and* $(0, \bar{u}) \in \partial_{(p,y)}[-L](x, p, y)$.

If $Ef(x,u)$ is closed in u, then $1 \Leftrightarrow 2 \Leftrightarrow 3$.

The subdifferential conditions in part 3 of 5.4 are known as the (generalized) *Karush-Kuhn-Tucker-Rockafellar* (KKTR) conditions; see Section 2.9.

In order to write the dual problem and the optimality conditions more explicitly in terms of the problem data, we will first derive explicit expressions for F^*. Section 5.1 will focus on the Lagrangian and the associated minimax problem.

Theorem 5.5. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$*, then* Ef^* *is closed,*

$$F^*(v,p,y) = Ef^*(v+p,y) + \delta_{\mathcal{X}_a^\perp}(p)$$

and, in particular,

$$\varphi^*(p,y) = Ef^*(p,y) + \delta_{\mathcal{X}_a^\perp}(p).$$

If, in addition, $\operatorname{dom} Ef^* \cap (\mathcal{V} \times \mathcal{Y}) \neq \emptyset$*, then the functions* Ef*,* Ef^**,* F *and* F^* *are all closed and proper.*

Proof. Recall that $F(x,z,u) = Ef(x,u) + \delta_{\mathcal{X}_a}(x-z)$, where $\mathcal{X}_a$ is closed by 5.1. By 4.8, the first assumption implies the closedness of Ef^* and, by the interchange rule in 3.29, that

$$\begin{aligned} F^*(v,p,y) &= \sup_{x\in\mathcal{X}, z\in\mathcal{X}, u\in\mathcal{U}} \{\langle x,v\rangle + \langle z,p\rangle + \langle u,y\rangle - Ef(x,u) \,|\, x-z \in \mathcal{X}_a\} \\ &= \sup_{x\in\mathcal{X}, z'\in\mathcal{X}, u\in\mathcal{U}} \{E[x\cdot(v+p) + u\cdot y - f(x,u) - z'\cdot p] \,|\, z' \in \mathcal{X}_a\} \\ &= Ef^*(v+p,y) + \delta_{\mathcal{X}_a^\perp}(p). \end{aligned}$$

The expression for φ^* now follows from the fact that $\varphi(p,y) = F^*(0,p,y)$, by definition. When $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ and $\operatorname{dom} Ef^* \cap (\mathcal{V} \times \mathcal{Y}) \neq \emptyset$, both Ef and Ef^* are closed and proper, by 4.9 and then, the functions F and F^* are closed as sums of closed functions. The properness of Ef and Ef^* clearly implies the properness of F and F^*. ☐

Corollary 5.6. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$*, the dual problem* (D) *can be written as*

$$\text{maximize} \quad \langle \bar{u}, y\rangle - Ef^*(p,y) \quad \text{over} \quad (p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$$

as well as

$$\text{maximize} \quad E[\bar{u}\cdot y - f^*(p,y)] \quad \text{over} \quad (p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$$

Proof. The first claim follows directly from 5.5. As to the second, Fenchel's inequality gives

$$f^*(p,y) \geq u\cdot y - f(x,u)$$

so the assumption $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ implies that the negative part of $f^*(p,y)$ is integrable for every $(p,y) \in \mathcal{V} \times \mathcal{Y}$. The claim thus follows from 3.33. ☐

The first condition in 5.5 clearly holds if $(P_{\mathcal{X}})$ is feasible. If in addition, the dual problem is feasible, then by 5.6, the second condition in 5.5 holds as well. Note that the dual is feasible e.g. if F is bounded from below since then $F^*(0,0)$ is finite.

In the deterministic setting, $\mathcal{X}_a^\perp = \{0\}$ so the dual problem becomes

$$\text{maximize} \quad \bar{u}\cdot y - f^*(0,y) \quad \text{over} \quad y \in \mathbb{R}^m$$

and we recover a finite-dimensional instance of the general conjugate duality framework; see Section 2.9. In general, the dual objective can be written also as

$$\langle \bar{u}, y\rangle - Ef^*(p,y) = E \inf_{(x,u)\in\mathbb{R}^n\times\mathbb{R}^m} [f(x,u) - x\cdot p + (\bar{u}-u)\cdot y].$$

This is the optimum value in a relaxed version of the primal problem (**??**) where we are now allowed to optimize over both x and u and the information constraint $x \in \mathcal{N}$ has been removed so the minimization can be done scenariowise; see 3.29. The constraints $x \in \mathcal{X}_a$ and $u = \bar{u}$ have been replaced by linear penalties given by the dual variables p and y.

Recall that the optimum value of (D) is always less than or equal to that of $(P_{\mathcal{X}})$. If the value function φ is closed at $(0,\bar{u})$ then, by 5.2, the optimum values are equal. If $(p,y) \in \partial\varphi(0,\bar{u})$ then, by 5.3, there is no duality gap and (p,y) solves the dual. This implies, in particular, that p is a subgradient of φ with respect to the first argument at $(0,\bar{u})$, i.e.

$$Ef(x+z,\bar{u}) - \langle z,p\rangle \geq \varphi(0,\bar{u}) \quad \forall x \in \mathcal{X}_a, z \in \mathcal{X}.$$

In other words, p describes a linear penalty that would make it disadvantageous to use nonadapted strategies in $(P_{\mathcal{X}})$. Such a $p \in \mathcal{X}_a^\perp$ is known as a *shadow price of information*.

When the dimension n_t of x_t and p_t is independent of time, the elements of $\mathcal{X}_a^\perp$ can be seen as nonadapted *martingale increments*. Indeed, we then have $p \in \mathcal{X}_a^\perp$ if and only if $p_t = \Delta m_{t+1}$ for $m_t \in \mathcal{V}_t$ and $m_{T+1} \in \mathcal{V}_T$ such that

$$E_t[\Delta m_{t+1}] = 0.$$

This is the usual martingale condition, but here m need not be adapted.

5.5 can be used to restate 5.2, 5.3 and 5.4 more explicitly. In particular, the first part of 5.4 can be written as follows.

Theorem 5.7. *If $(P_{\mathcal{X}})$ and (D) are feasible, then the following are equivalent:*

1. *x solves $(P_{\mathcal{X}})$, (p,y) solves (D) and* $\inf(P_{\mathcal{X}}) = \sup(D)$*;*
2. *$x \in \mathcal{X}_a$, $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ and*

$$(p,y) \in \partial f(x,\bar{u}) \quad a.s.$$

Proof. By 5.4, 1 is equivalent to $(0, p, y) \in \partial F(x, 0, \bar{u})$ which means that $F(x, 0, \bar{u}) + F^*(0, p, y) = \langle \bar{u}, y \rangle$. By Lemma 5.5, this means that $x \in \mathcal{X}_a$, $p \in \mathcal{X}_a^\perp$ and

$$Ef(x, \bar{u}) + Ef^*(p, y) = E[x \cdot p] + E[\bar{u} \cdot y]. \tag{5.1}$$

Given $(x', u') \in \mathcal{X} \times \mathcal{U}$ and $(p', y') \in \mathcal{V} \times \mathcal{Y}$, we have

$$f(x', u') + f^*(p', y') \geq x' \cdot p' + u \cdot y', \tag{5.2}$$

by Fenchel's inequality, so the feasibility assumptions imply that the negative parts of $f(x', u')$ and $f^*(p', y')$ are integrable. Thus, by 3.33,

$$Ef(x', u') + Ef^*(p', y') = E[f(x', u') + f^*(p', y')]$$

so (5.1) means that $(x, \bar{u})$ and (p, y) satisfy (5.2) as an equality, i.e. $(p, y) \in \partial f(x, \bar{u})$. □

If the subdifferential $\partial\varphi(0, \bar{u})$ is nonempty, then by 5.3, there is no duality gap and a dual has a solution. 5.7 thus implies the following.

Corollary 5.8. *If $\partial\varphi(0, \bar{u}) \neq \emptyset$, then $\inf(P_{\mathcal{X}}) = \sup(D)$, the dual optimum is attained and the following are equivalent for an $x \in \mathcal{X}_a$:*

1. *x solves $(P_{\mathcal{X}})$;*
2. *there exists $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ with*

$$(p, y) \in \partial f(x, \bar{u}) \quad a.s.$$

5.1 Lagrangian integrands and KKTR-conditions

This section focuses on the Lagrangian L and the associated minimax problem. The Lagrangian L itself has a somewhat cumbersome expression but it turns out that it is "equivalent" to a simpler function that has the same saddle value and saddle points. The expressions derived below, involve the *Lagrangian integrand* $l : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ defined by

$$l(x, y, \omega) := \inf_{u \in \mathbb{R}^m} \{f(x, u, \omega) - u \cdot y\}.$$

For any (x, y, ω), the function $l(\cdot, y, \omega)$ is convex, by 2.8, and $l(x, \cdot, \omega)$ is upper semicontinuous and concave. Clearly,

$$f^*(v, y, \omega) = \sup_{x \in \mathbb{R}^n} \{x \cdot v - l(x, y, \omega)\}$$

so, by 2.23,

$$(\mathrm{cl}_x\, l)(x, y, \omega) = \sup_{v \in \mathbb{R}^n} \{x \cdot v - f^*(v, y, \omega)\},$$

where, for each $(y,\omega)\in\mathbb{R}^m\times\Omega$, the function $(\mathrm{cl}_x\, l)(\cdot,y,\omega)$ denotes the *closure* of the function $l(\cdot,y,\omega)$; see Section **??**.

Given $x\in\mathcal{X}$, the function

$$(y,\omega)\mapsto -l(x(\omega),y,\omega)=\sup_{u\in\mathbb{R}^m}\{u\cdot y-f(x(\omega),u,\omega)\}$$

is a normal integrand, by 3.9 and 3.26. Similarly, the function

$$(x,\omega)\mapsto (\mathrm{cl}_x\, l)(x,y(\omega),\omega)=\sup_{v\in\mathbb{R}^n}\{x\cdot v-f^*(v,y(\omega),\omega)\}$$

is normal integrand for any $y\in\mathcal{Y}$. Thus, by **??**, the functions

$$\omega\mapsto l(x(\omega),y(\omega),\omega)\quad\text{and}\quad \omega\mapsto(\mathrm{cl}_x\, l)(x(\omega),y(\omega),\omega)$$

are measurable for any $(x,y)\in\mathcal{X}\times\mathcal{Y}$. It follows that the integral functionals

$$El(x,y):=\int_\Omega l(x(\omega),y(\omega),\omega)dP(\omega)$$

and

$$E(\mathrm{cl}_x\, l)(x,y):=\int_\Omega (\mathrm{cl}_x\, l)(x(\omega),y(\omega),\omega)dP(\omega)$$

are well-defined extended real-valued functions on $\mathcal{X}\times\mathcal{Y}$.

We will denote the projection of $\mathrm{dom}\, Ef$ to the x-component by

$$\mathrm{dom}_x\, Ef:=\{x\in\mathcal{X}\mid \exists u\in\mathcal{U}:\ Ef(x,u)<\infty\}$$

and the projection of $\mathrm{dom}\, Ef^*$ to the y-component by

$$\mathrm{dom}_y\, Ef^*:=\{y\in\mathcal{Y}\mid \exists v\in\mathcal{U}:\ Ef^*(v,y)<\infty\}.$$

Theorem 5.9. *We have*

$$L(x,p,y)=\begin{cases}+\infty & \textit{if } x\notin\mathrm{dom}_x\, Ef,\\ El(x,y)-\langle x,p\rangle & \textit{if } x\in\mathrm{dom}_x\, Ef \textit{ and } p\in\mathcal{X}_a^\perp,\\ -\infty & \textit{otherwise.}\end{cases}$$

If $\mathrm{dom}\, Ef\cap(\mathcal{X}\times\mathcal{U})\neq\emptyset$, *then*

$$(\mathrm{cl}_x\, L)(x,p,y)=\begin{cases}E(\mathrm{cl}_x\, l)(x,y)-\langle x,p\rangle & \textit{if } y\in\mathrm{dom}_y\, Ef^* \textit{ and } p\in\mathcal{X}_a^\perp,\\ -\infty & \textit{otherwise.}\end{cases}$$

If $\mathrm{dom}\, Ef\cap(\mathcal{X}\times\mathcal{U})\neq\emptyset$ *and* $\mathrm{dom}\, Ef^*\cap(\mathcal{V}\times\mathcal{Y})\neq\emptyset$, *then all functions between* L *and* $\mathrm{cl}_x L$ *have the same saddle value and saddle points. In this case, the KKTR-conditions*

$$0\in\partial_x L(x,p,y),\quad (0,\bar u)\in\partial_{p,y}[-L](x,p,y)$$

in 5.4 hold if and only if $x\in\mathcal{X}_a$, $p\in\mathcal{X}_a^\perp$ *and*

$$p\in\partial_x l(x,y),\quad \bar u\in\partial_y[-l](x,y)\quad \textit{a.s.}$$

Proof. By 3.33,

$$\begin{aligned}L(x,p,y) &= \inf_{(z,u)\in\mathcal{X}\times\mathcal{U}}\{F(x,z,u)-\langle z,p\rangle-\langle u,y\rangle\}\\ &= \inf_{(z,u)\in\mathcal{X}\times\mathcal{U}}\{E[f(x,u)-z\cdot p-u\cdot y]\,|\,x-z\in\mathcal{X}_a\}\\ &= \inf_{(z',u)\in\mathcal{X}\times\mathcal{U}}\{E[f(x,u)-(x-z')\cdot p-u\cdot y]\,|\,z'\in\mathcal{X}_a\},\end{aligned}$$

so the expression for L follows from 3.29. By 5.5 and 3.33,

$$\begin{aligned}(\mathrm{cl}_x\,L)(x,p,y) &= \sup_{v\in\mathcal{V}}\{\langle x,v\rangle-F^*(v,p,y)\}\\ &= \begin{cases}\sup_{v\in\mathcal{V}}\{\langle x,v\rangle-Ef^*(v+p,y)\} & \text{if } p\in\mathcal{X}_a^\perp,\\ -\infty & \text{otherwise}\end{cases}\\ &= \begin{cases}\sup_{v\in\mathcal{V}}E[x\cdot v-f^*(v+p,y)] & \text{if } p\in\mathcal{X}_a^\perp,\\ -\infty & \text{otherwise}\end{cases}\end{aligned}$$

so the expression for $\mathrm{cl}_x\,L$ follows from 3.29 again. When $\mathrm{dom}\,Ef\neq\emptyset$ and $\mathrm{dom}\,Ef^*\neq\emptyset$, the function F is closed and proper, by 5.5, so the claims about saddle value and saddle points follow from 2.37.

By 5.4, the KKTR-conditions hold if and only if $(0,p,y)\in\partial F(x,0,\bar u)$ or, equivalently, if

$$F(x,0,\bar u)+F^*(0,p,y)=\langle\bar u,y\rangle.$$

By 5.5, this means that $x\in\mathcal{X}_a$, $p\in\mathcal{X}_a^\perp$ and

$$Ef(x,\bar u)+Ef^*(p,y)=E[\bar u\cdot y]$$

or, equivalently,

$$Ef(x,\bar u)+Ef^*(p,y)=E[x\cdot p]+E[\bar u\cdot y].$$

Since, by Fenchel's inequality,

$$f(x,\bar u,\omega)+f^*(v,y,\omega)\geq x\cdot v+\bar u\cdot y,$$

this means that $(p,y)\in\partial f(x,\bar u)$ almost surely. By 2.33, this is equivalent to $v\in\partial_x l(x,y)$ and $\bar u\in\partial_y[-l](x,y)$. □

The functions L and $\mathrm{cl}_x\,L$ are not quite integral functionals because of the constraints on the variables. However, one of the saddle functions between L and $\mathrm{cl}_x\,L$ is the function

$$\tilde L(x,p,y)=\begin{cases}E[l(x,y)-x\cdot p] & \text{if } p\in\mathcal{X}_a^\perp,\\ -\infty & \text{otherwise.}\end{cases}$$

5.4, 5.9 and 2.37 thus yield the following extension of 5.7.

Corollary 5.10. *If $(P_{\mathcal{X}})$ and (D) are feasible, the following are equivalent:*

1. *x solves $(P_{\mathcal{X}})$, (p,y) solves (D) and* $\inf(P_{\mathcal{X}}) = \sup(D)$*;*
2. *$x \in \mathcal{X}_a$, $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ and*
$$(p,y) \in \partial f(x,\bar{u}) \quad a.s.;$$
3. *(x,p,y) is a saddle point of the integral functional*
$$(x,p,y) \mapsto E[l(x,y) - x \cdot p + \bar{u} \cdot y],$$
when minimizing over $x \in \mathcal{X}$ and maximizing over $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$;
4. *$x \in \mathcal{X}_a$, $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ and*
$$p \in \partial_x l(x,y), \quad \bar{u} \in \partial_y[-l](x,y) \quad a.s.$$

Similarly, we can augment 5.8 as follows.

Corollary 5.11. *If $\partial\varphi(0,\bar{u}) \neq \emptyset$, the following are equivalent for an $x \in \mathcal{X}_a$:*

1. *x solves $(P_{\mathcal{X}})$;*
2. *there exists $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ with*
$$(p,y) \in \partial f(x,\bar{u}) \quad a.s.;$$
3. *there exists $(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ with*
$$p \in \partial_x l(x,y), \quad \bar{u} \in \partial_y[-l](x,y) \quad a.s.$$

In the deterministic setting, $\mathcal{X}_a^\perp = \{0\}$ so condition 3 in 5.11 becomes the KKTR-condition in finite-dimensional convex optimization. In the stochastic setting, the shadow price of information $p \in \mathcal{X}_a^\perp$ allows us to write the KKTR-conditions scenariowise.

5.2 Reduced dual problems

In many applications, one can restrict the dual variables (p,y) to a subset of $\mathcal{X}_a^\perp \times \mathcal{Y}$ without lowering the optimum value of the dual problem. This happens, in particular, under the following.

Assumption 5.12. *There is a mapping $\Pi : \mathcal{V} \times \mathcal{Y} \to \mathcal{V} \times \mathcal{Y}$ such that*
$$\varphi^* \circ \Pi \leq \varphi^*$$
and $\langle (0,\bar{u}), \Pi(p,y) \rangle = \langle \bar{u}, y \rangle$ for all $(p,y) \in \operatorname{dom} \varphi^*$.

Indeed, under 5.12 the optimum value of (D) equals that of the problem

$$\text{maximize} \quad \langle \bar{u}, y\rangle - \varphi^*(p, y) \quad \text{over}(p, y) \in \operatorname{rge}\Pi \tag{5.3}$$

while

$$\Pi(\operatorname{argmax}(D)) \subseteq \operatorname{argmax}(5.3) = \operatorname{argmax}(D) \cap \operatorname{rge}\Pi.$$

5.12 is clearly satisfied if Π is the identity mapping but in many situations, more interesting choices are available.

By 2.8, the function

$$g(y) := \inf_{p\in\mathcal{V}} \varphi^*(p, y)$$

is convex on $\mathcal{Y}$. It is clear that the optimum value of (D) equals that of the problem

$$\frac{\text{maximize} \quad \langle \bar{u}, y\rangle - g(y)}{y \in \mathcal{Y}} \tag{5.4}$$

and that a pair (p, y) solves (D) if and only if y solves (5.4) and p attains the infimum in the definition of g. Recall that a mapping $\gamma : \mathcal{Y} \to \mathcal{Y}$ is *idempotent* if $\gamma\circ\gamma = \gamma$.

Theorem 5.13 (Reduced dual)**.** *Assume that there exist mappings $\pi : \mathcal{Y} \to \mathcal{V}$ and $\gamma : \mathcal{Y} \to \mathcal{Y}$ such that γ is idempotent and the mapping $\Pi(p, y) = (\pi(y), \gamma(y))$ satisfies 5.12. Then*

$$g(y) = \varphi^*(\pi(y), y) \quad \forall y \in \operatorname{rge}\gamma,$$

optimum value of (D) coincides with that of the problem

$$\frac{\text{maximize} \quad \langle \bar{u}, y\rangle - g(y)}{y \in \operatorname{rge}\gamma,} \tag{5.5}$$

and if $y \in \operatorname{rge}\gamma$ solves (5.5) then $(\pi(y), y)$ solves (D). If (p, y) solves (D), then $\gamma(y)$ solves (5.5). If $(P_{\mathcal{X}})$ and (5.5) are feasible, then the following are equivalent:

1. *x solves $(P_{\mathcal{X}})$, y solves (5.5) and $\inf(P_{\mathcal{X}}) = \sup(5.5)$;*
2. *$x \in \mathcal{X}_a$, $y \in \operatorname{rge}\gamma$ and*

$$(\pi(y), y) \in \partial f(x, \bar{u}) \quad a.s.;$$

3. *$x \in \mathcal{X}_a$, $y \in \operatorname{rge}\gamma$ and*

$$\pi(y) \in \partial_x l(x, y), \quad \bar{u} \in \partial_y[-l](x, y) \quad a.s.$$

Proof. Given $y \in \mathcal{Y}$, we have

$$\begin{aligned} g(\gamma(y)) &= \inf_{p \in \mathcal{V}} \varphi^*(p, \gamma(y)) \\ &\le \varphi^*(\pi(y), \gamma(y)) \\ &\le \inf_{p \in \mathcal{V}} \varphi^*(p, y) \\ &= g(y), \end{aligned}$$

where the second inequality holds by the assumption on Π. Combining the above with $\langle (0, \bar{u}), \Pi(p, y) \rangle = \langle \bar{u}, y \rangle$, we get

$$\sup (5.5) \ge \sup (5.3) \ge \sup (5.4),$$

where equalities must hold since, trivially, $\sup (5.4) \ge \sup (5.5)$. Thus, the optimum value of (5.5) equals that of (5.4) which in turn equals $\sup (D)$. When γ is idempotent and $y \in \operatorname{rge} \gamma$, we have $\gamma(y) = y$ so the above hold with equalities and

$$g(y) = \varphi^*(\pi(y), y).$$

This gives the relations between the optimal solutions. The rest now follows from 5.10. □

The conditions of 5.13 may seem rather special, but they are satisfied in many applications; see Section 7 for examples. In the applications, the mappings π and γ are typically defined in terms of conditional expectations.

Theorem 5.14. *If g is closed at $y \in \mathcal{Y}$ and φ is closed at $(0, u)$ for all $u \in \mathcal{U}$, then $g(y) = \varphi(0, \cdot)^*(y)$. In particular, if g and φ are closed, then $g = \varphi(0, \cdot)^*$.*

Proof. By 2.23,

$$\begin{aligned} g^*(u) &= \sup_{y} \{\langle u, y \rangle - g(y)\} \\ &= \sup_{p,y} \{\langle u, y \rangle - \varphi^*(p, y)\} \\ &= (\operatorname{cl} \varphi)(0, u). \end{aligned}$$

Under the closedness assumptions, another application of 2.23 proves the claim. □

Remark 5.15. *Under the assumptions of 5.14, the reduced dual* (5.4) *is the dual problem obtained from the general conjugate duality but without the parameter $z \in \mathcal{X}$. Unlike φ^*, however, the function g is not closed, in general.*

The function g is closed under the assumptions of 5.13 if π is continuous and γ is the identity mapping. Indeed, 5.13 then says that $g(y) = \varphi^*(\pi(y), y)$ which is closed as a composition of a continuous linear mapping and a closed convex function.

The following lemma gives sufficient conditions for 5.12.

Lemma 5.16. *Let* $\Pi : \mathcal{V}\times\mathcal{Y}\to\mathcal{V}\times\mathcal{Y}$ *and* $\xi : \mathcal{X}\times\mathcal{X}\times\mathcal{U}\to\mathcal{X}$ *be such that* Π *is continuous and linear and*

$$F(\xi(x,z,u),\Pi^*(z,u)) \le F(x,z,u) \quad \forall (x,z,u)\in\mathcal{X}\times\mathcal{X}\times\mathcal{U}.$$

Then

$$\varphi\circ\Pi^* \le \varphi \quad \textit{and} \quad \varphi^*\circ\Pi \le \varphi^*.$$

If, in addition, $\Pi^*(0,\bar{u})\in\mathcal{X}_a\times\{\bar{u}\}$*, then 5.12 holds.*

Proof. Minimizing both sides of the inequality over $x\in\mathcal{X}$ gives

$$\inf_x F(x,z,u) \le \inf_x F(\xi(x,z,u),\Pi^*(z,u)) \le \inf_x F(x,z,u) \quad \forall (z,u)\in\mathcal{X}\times\mathcal{U}$$

or, in other words, $\varphi\circ\Pi^*\le\varphi$. By **??**, this implies $\varphi^*\circ\Pi\le\varphi^*$.

Assume now that $\Pi^*(0,\bar{u})\in\mathcal{X}_a\times\{\bar{u}\}$. Since $\operatorname{dom}\varphi^*\subseteq\mathcal{X}_a^\perp\times\mathcal{Y}$, we get

$$\langle(0,\bar{u}),\Pi(p,y)\rangle = \langle\Pi^*(0,\bar{u}),(p,y)\rangle = \langle\bar{u},y\rangle$$

for all $(p,y)\in\operatorname{dom}\varphi^*$. □

6 Duality for (P)

We now return to problem (**??**) where one optimizes over the space $\mathcal{N}$ of all adapted strategies, not just those belonging to $\mathcal{X}$ as in $(P_\mathcal{X})$. While problem $(P_\mathcal{X})$ in Section 5 allows for a convenient dualization within the purely functional analytic conjugate duality framework, there are interesting applications where $\inf(P_\mathcal{X}) > \inf$ (**??**) or where the infimum in (**??**) is attained in L^0 but not in $\mathcal{X}$; see 6.6 for a simple illustration. It may even happen that $(P_\mathcal{X})$ is infeasible while (**??**) is not. This section shows that many of the duality relations between $(P_\mathcal{X})$ and (D) derived in Section 5 also hold between (**??**) and (D).

Recall the dual of $(P_\mathcal{X})$ can be written as

$$\text{maximize} \quad \langle\bar{u},y\rangle-\varphi^*(p,y) \quad \text{over}(p,y)\in\mathcal{V}\times\mathcal{Y}. \tag{D}$$

One could define a dual problem for (**??**) simply by replacing φ^* by the conjugate of the function $\bar{\varphi}:\mathcal{X}\times\mathcal{U}\to\overline{\mathbb{R}}$ defined by

$$\bar{\varphi}(z,u) := \inf_{x\in L^0}\{Ef(x,u)\mid x-z\in\mathcal{N}\}.$$

While in the definition of φ, the strategies are sought from the locally convex space $\mathcal{X}\subset L^0$, in the definition of $\bar{\varphi}$, we minimize over all of L^0. Clearly, $\bar{\varphi}(0,\bar{u}) = \inf$ (**??**) and $\bar{\varphi}\le\varphi$. Under a mild condition, the conjugates of φ and $\bar{\varphi}$ coincide, so we may regard (D) as the dual problem of both $(P_\mathcal{X})$ and (**??**).

Lemma 6.1. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$*, then* $\varphi^* = \bar{\varphi}^*$ *and*

$$\partial\varphi(z,u) \subseteq \partial\bar{\varphi}(z,u)$$

for every $(z,u) \in \mathcal{X} \times \mathcal{U}$ *with an equality whenever the left side is nonempty.*

Proof. Since $\varphi \geq \bar{\varphi}$, we have $\varphi^* \leq \bar{\varphi}^*$. To prove the converse, let $(p,y) \in \operatorname{dom}\varphi^*$. By 5.5,

$$\varphi^*(p,y) = Ef^*(p,y) + \delta_{\mathcal{X}_a^\perp}(p),$$

so $p \in \mathcal{X}_a^\perp$. Given any $(x,z,u) \in L^0 \times \mathcal{X} \times \mathcal{U}$, Fenchel's inequality gives

$$Ef(x,u) + \delta_{\mathcal{N}}(x-z) + Ef^*(p,y) \geq E[(x-z)\cdot p] + E[z\cdot p] + E[u\cdot y]$$

so, by 3.40,

$$Ef(x,u) + \delta_{\mathcal{N}}(x-z) + Ef^*(p,y) \geq E[z\cdot p] + E[u\cdot y].$$

Thus, $\bar{\varphi}(z,u) + \varphi^*(p,y) \geq \langle z,p\rangle + \langle u,y\rangle$ for all $(z,u) \in \mathcal{X}\times\mathcal{U}$, which means that $\bar{\varphi}^*(p,y) \leq \varphi^*(p,y)$. This proves the first claim.

Trivially, $\partial\varphi(z,u) \subseteq \partial\bar{\varphi}(z,u)$ if $\partial\varphi(z,u) = \emptyset$, so assume that $\partial\varphi(z,u) \neq \emptyset$. We have, in particular, $(z,u) \in \operatorname{dom}\varphi$ and thus, $\operatorname{dom} Ef \cap (\mathcal{X}\times\mathcal{U}) \neq \emptyset$. By the first claim, $\varphi^* = \bar{\varphi}^*$. Recall that, by Fenchel's inequality,

$$\varphi(z,u) + \varphi^*(p,y) \geq \langle z,p\rangle + \langle u,y\rangle$$

and that the equality holds if and only if $(p,y) \in \partial\varphi(z,u)$. Similarly for $\bar{\varphi}$. The subdifferential inclusion thus follows from the fact that $\varphi \geq \bar{\varphi}$.

By the first claim and 2.23, $\operatorname{cl}\varphi = \operatorname{cl}\bar{\varphi}$. In particular, $\varphi \geq \bar{\varphi} \geq \operatorname{cl}\varphi$. When $\partial\varphi(z,u) \neq \emptyset$, we have $\varphi(z,u) = \operatorname{cl}\varphi(z,u)$ so $\bar{\varphi}(z,u) = \varphi(z,u)$ and thus, $\bar{\varphi}(z,u) + \bar{\varphi}^*(p,y) = \langle x,p\rangle + \langle u,y\rangle$ if and only if $\varphi(z,u) + \varphi^*(p,y) = \langle x,p\rangle + \langle u,y\rangle$. In other words, $(p,y) \in \partial\bar{\varphi}(z,u)$ if and only if $(p,y) \in \partial\varphi(z,u)$. □

The following summarizes the relationships between problems $(P_{\mathcal{X}})$, (**??**) and (D).

Theorem 6.2. *Assume that* $\operatorname{dom} Ef \cap (\mathcal{X}\times\mathcal{U}) \neq \emptyset$*. We have*

$$\inf(P_{\mathcal{X}}) \geq \inf(\textbf{??}) \geq \sup(D)$$

and

1. $\inf(\textbf{??}) = \sup(D)$ *if and only if* $\bar{\varphi}$ *is closed at* $(0,\bar{u})$*,*
2. $\inf(P_{\mathcal{X}}) = \inf(\textbf{??}) = \sup(D)$ *if and only if* φ *is closed at* $(0,\bar{u})$*,*
3. *if* $\bar{\varphi}(0,\bar{u}) < \infty$*, then the following are equivalent:*
   (a) (p,y) *solves* (D) *and* $\inf(\textbf{??}) = \sup(D)$*;*

(b) *either* $(p, y) \in \partial\bar{\varphi}(0, \bar{u})$ *or* $\bar{\varphi}(0, \bar{u}) = -\infty$,

4. *if* $\varphi(0, \bar{u}) < \infty$, *then the following are equivalent:*

   (a) (p, y) *solves* (D) *and* $\inf(P_{\mathcal{X}}) = \inf(\textbf{??}) = \sup(D)$;

   (b) *either* $(p, y) \in \partial\varphi(0, \bar{u})$ *or* $\varphi(0, \bar{u}) = -\infty$.

Proof. The first inequality is trivial. Fenchel's inequality gives

$$\begin{aligned}\inf(\textbf{??}) &= \bar{\varphi}(0, \bar{u}) \\ &\geq \sup_{(p,y)\in\mathcal{V}\times\mathcal{Y}} \{\langle \bar{u}, y\rangle - \bar{\varphi}^*(p, y)\} \\ &= \sup_{(p,y)\in\mathcal{V}\times\mathcal{Y}} \{\langle \bar{u}, y\rangle - \varphi^*(p, y)\} = \sup(D),\end{aligned}$$

where the second equality holds by 6.1.

By definition, $\inf(\textbf{??}) = \bar{\varphi}(0, \bar{u})$ and $\sup(D) = \varphi^{**}(0, \bar{u})$. By 6.1, $\varphi^{**} = \bar{\varphi}^{**}$ so part 1 follows from 2.23. Part 2 follows from 5.2 while part 4 follows from 5.3. It remains to prove 3. By Fenchel's inequality,

$$\bar{\varphi}(0, z) \geq \langle \bar{u}, y\rangle - \bar{\varphi}^*(p, y).$$

Condition 3a means that either $\bar{\varphi}(0, \bar{u}) = -\infty$ or $\bar{\varphi}(0, z) = \langle \bar{u}, y\rangle - \varphi^*(p, y)$ while, by the definition of a subgradient, 3b means that either $\bar{\varphi}(0, \bar{u}) = -\infty$ or $\bar{\varphi}(0, z) = \langle \bar{u}, y\rangle - \bar{\varphi}^*(p, y)$. The claim thus follows from 6.1. □

The condition $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ in 6.1 and 6.2 holds, in particular, if $(P_{\mathcal{X}})$ is feasible. Sufficient conditions for the closedness of $\bar{\varphi}$ will be given in Chapter **??** while Chapter **??** gives sufficient conditions for the subdifferentiability of φ at $(0, \bar{u})$.

The following gives an analogue of 5.10 for general strategies $x \in L^0$.

Theorem 6.3. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (**??**) *and* (D) *are feasible, then the following are equivalent:*

1. x *solves* (**??**), (p, y) *solves* (D) *and* $\inf(\textbf{??}) = \sup(D)$;

2. x *is feasible in* (**??**), (p, y) *is feasible in* (D) *and*

$$(p, y) \in \partial f(x, \bar{u}) \quad a.s.;$$

3. x *is feasible in* (**??**), (p, y) *is feasible in* (D) *and*

$$p \in \partial_x l(x, y), \quad \bar{u} \in \partial_y[-l](x, y) \quad a.s.$$

Proof. The equivalence of 2 and 3 follows by scenariowise application of the equivalence of 3 and 5 in 2.33. Let $x \in \mathcal{N}$ and $(p, y) \in \mathcal{V} \times \mathcal{Y}$ be feasible. By Fenchel's inequality,

$$f(x, \bar{u}) + f^*(p, y) - \bar{u} \cdot y \geq x \cdot p \quad a.s. \tag{6.1}$$

so, by 3.33,

$$Ef(x, \bar{u}) + E[f^*(p, y) - \bar{u} \cdot y] \geq E[x \cdot p]. \tag{6.2}$$

By the feasibility of x and (p, y), the expectations on the left are finite so (6.2) holds as an equality if and only if (6.1) holds as an equality almost surely. Equality in (6.1) means that 2 holds. By Lemma 3.40, $E[x \cdot p] = 0$, so equality in (6.2) means that 1 holds. □

If $\partial\bar{\varphi}(0, \bar{u}) \neq \emptyset$, then, by 6.2, inf (**??**) = sup (D) and the dual has a solution. 6.3 thus implies the following optimality conditions for (**??**).

Corollary 6.4. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* $\partial\bar{\varphi}(0, \bar{u}) \neq \emptyset$*, then* inf (**??**) = sup (D)*, the dual optimum is attained, and the following are equivalent:*

1. *x solves* (**??**)*;*
2. *x is feasible in* (**??**) *and there exists (p, y) feasible in (D) with*

$$(p, y) \in \partial f(x, \bar{u}) \quad a.s.;$$

3. *x is feasible in* (**??**) *and there exists (p, y) feasible in (D) with*

$$p \in \partial_x l(x, y), \quad \bar{u} \in \partial_y[-l](x, y) \quad a.s.$$

Recall that, by 6.1, the condition $\partial\bar{\varphi}(0, \bar{u}) \neq \emptyset$ is implied by $\partial\varphi(0, \bar{u}) \neq \emptyset$. Sufficient conditions for this will be given in Chapter **??**.

Corollary 6.5. *Assume that* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* $(p, y) \in \partial\bar{\varphi}(0, \bar{u})$*. Then optimal solutions x of* (**??**) *are scenariowise minimizers of the function*

$$x \mapsto l(x, y(\omega), \omega) - x \cdot p(\omega).$$

Conversely, if the function has a unique scenariowise minimizer x and if (**??**) *admits solutions, then x solves* (**??**) *and, in particular, x is feasible and adapted.*

Proof. The first claim follows directly from 6.4 after observing that the first inclusion in part 3 means that x minimizes the function

$$x \mapsto l(x, y(\omega), \omega) - x \cdot p(\omega).$$

If the primal admits solutions, then it satisfies the inclusions in part 3 of 6.4. If the minimizer x of the above function is unique, it thus has to be a unique primal solution. □

Example 6.6. *It may happen that*

$$\inf(P_{\mathcal{X}}) > \inf(\textbf{??}) = \sup(D).$$

Indeed, let

$$f(x,u,\omega) = \delta_{\{0\}}(x_T - u\xi(\omega)).$$

If $\xi \notin \mathcal{X}$ *and* $\bar{u} = 1$*, then* (**??**) *is feasible while* $(P_{\mathcal{X}})$ *is not. Clearly* Ef *is proper on* $\mathcal{X} \times \mathcal{U}$ *and* $f^*(0,0) = 0$*, so* $\inf(\textbf{??}) = \sup(D) = 0$*. Another example with finite* $\inf(P_{\mathcal{X}}) < \infty$ *is obtained by letting*

$$f(x,u,\omega) = (x_0 - 1)^2 + \delta_{\{0\}}(x_0\xi(\omega) - x_1),$$

$\mathcal{F}_0 = \{\Omega, \emptyset\}$ *and* $\xi \in L^0(\mathcal{F}_1)$ *with* $\xi \notin \mathcal{X}$*. Since* f *is nonnegative,* $(1,\xi)$ *is optimal for* (**??**) *and the optimum value is zero. Here* Ef *is proper on* $\mathcal{X} \times \mathcal{U}$*, and, by a direct verification,* $f^*(0,0) = 0$*, so the origin is a dual solution and* $\inf(\textbf{??}) = \sup(D) = 0$*. On the other hand, the only feasible solution of* $(P_{\mathcal{X}})$ *is the origin, so* $\inf(P_{\mathcal{X}}) = 1$*.*

Recall the reduced dual problems and the function

$$g(y) := \inf_{p\in\mathcal{V}} \varphi^*(p,y)$$

from Section 5.2. The following extends 5.13 to general strategies $x \in \mathcal{N}$.

Theorem 6.7 (Reduced dual)**.** *Assume that there exist mappings* $\pi : \mathcal{Y} \to \mathcal{V}$ *and* $\gamma : \mathcal{Y} \to \mathcal{Y}$ *such that* γ *is idempotent and the mapping* $\Pi(p,y) = (\pi(y), \gamma(y))$ *satisfies 5.12. Then*

$$g(y) = \varphi^*(\pi(y), y) \quad \forall y \in \operatorname{rge}\gamma,$$

optimum value of (D) *coincides with that of the problem*

$$\begin{array}{c}\text{maximize} \quad \langle \bar{u}, y\rangle - g(y) \\ \hline y \in \operatorname{rge}\gamma,\end{array} \tag{6.3}$$

and if $y \in \operatorname{rge}\gamma$ *solves* (6.3) *then* $(\pi(y), y)$ *solves* (D)*. If* (p,y) *solves* (D)*, then* $\gamma(y)$ *solves* (6.3)*. If* (**??**) *and* (6.3) *are feasible, then the following are equivalent:*

1. x *solves* (**??**), y *solves* (6.3) *and* $\inf(\textbf{??}) = \sup(6.3)$*;*
2. x *is feasible in* (**??**), y *is feasible in* (6.3) *and*

$$(\pi(y), y) \in \partial f(x, \bar{u}) \quad a.s.;$$

3. x *is feasible in* (**??**), y *is feasible in* (6.3) *and*

$$\pi(y) \in \partial_x l(x,y), \quad \bar{u} \in \partial_y[-l](x,y) \quad a.s.$$

Proof. The claims up to the equivalences are a repetition of 5.13. The equivalences follow now from 6.3. □

The following is analogous to 5.14.

Theorem 6.8. *Assume that* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$. *If g is closed at $y \in \mathcal{Y}$ and $\bar{\varphi}$ is closed at $(0,u)$ for all $u \in \mathcal{U}$, then $g(y) = \bar{\varphi}(0,\cdot)^*(y)$. In particular, if g and $\bar{\varphi}$ are closed, then $g = \bar{\varphi}(0,\cdot)^*$.*

Proof. By 2.23 and 6.1,

$$\begin{aligned} g^*(u) &= \sup_y \{\langle u,y\rangle - g(y)\} \\ &= \sup_{p,y} \{\langle u,y\rangle - \varphi^*(p,y)\} \\ &= (\operatorname{cl}\varphi)(0,u) = (\operatorname{cl}\bar{\varphi})(0,u). \end{aligned}$$

Under the closedness assumptions, another application of 2.23 proves the claim. □

The closedness of $\bar{\varphi}$ will be the topic of the Chapter **??**. The following is the analogue of 5.16 for general strategies x. The proof is almost identical so it is omitted.

Lemma 6.9. *Let $\Pi : \mathcal{V} \times \mathcal{Y} \to \mathcal{V} \times \mathcal{Y}$ and $\xi : L^0 \times \mathcal{X} \times \mathcal{U} \to L^0$ be such that Π is continuous and linear and*

$$F(\xi(x,z,u), \Pi^*(z,u)) \leq F(x,z,u) \quad \forall (x,z,u) \in L^0 \times \mathcal{X} \times \mathcal{U}.$$

Then

$$\bar{\varphi} \circ \Pi^* \leq \bar{\varphi} \quad \text{and} \quad \bar{\varphi}^* \circ \Pi \leq \bar{\varphi}^*.$$

If, in addition, $\Pi^*(0,\bar{u}) \in \mathcal{X}_a \times \{\bar{u}\}$, then 5.12 holds.

7 Applications

This section applies the general duality results of this chapter to the five examples considered at the beginning of the course. We find explicit expressions for the involved functions and conditions but only give selected statements as examples of how the general results can be applied.

7.1 Mathematical programming

Consider again problem

$$\begin{aligned} &\text{minimize} && Ef_0(x) \quad \text{over } x \in \mathcal{N}, \\ &\text{subject to} && f_j(x) \leq 0 \quad j = 1,\dots,l \ a.s., \\ &&& f_j(x) = 0 \quad j = l+1,\dots,m \ a.s. \end{aligned}$$

This fits the general duality framework with $\bar{u} = 0$ and

$$f(x,u,\omega) = \begin{cases} f_0(x,\omega) & \text{if } x \in \operatorname{dom} H(\cdot,\omega),\ H(x,\omega)+u \in K, \\ +\infty & \text{otherwise,} \end{cases}$$

where $K = \mathbb{R}^l_- \times \{0\}$ and H is the random K-convex function defined by

$$\operatorname{dom} H(\cdot,\omega) = \bigcap_{j=1}^m \operatorname{dom} f_j(\cdot,\omega) \quad \text{and} \quad H(x,\omega) = (f_i(x,\omega))_{j=1}^m.$$

That f is a convex normal integrand follows from arguments similar to those used in Section **??** to show that h is a convex normal integrand.

The Lagrangian integrand becomes

$$\begin{aligned} l(x,y,\omega) &= \inf\{f(x,u,\omega) - u \cdot y\} \\ &= \inf\{f_0(x,\omega) - u \cdot y \mid x \in \operatorname{dom} H(\cdot,\omega),\ H(x,\omega)+u \in K\} \\ &= \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} H(\cdot,\omega), \\ f_0(x,\omega) + y \cdot H(x,\omega) & \text{if } x \in \operatorname{dom} H(\cdot,\omega) \text{ and } y \in K^\circ, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The conjugate of f is given by

$$\begin{aligned} f^*(p,y,\omega) &= \sup_{x\in\mathbb{R}^n} \{x \cdot p - l(x,y,\omega)\} \\ &= \sup_{x\in\mathbb{R}^n} \{x \cdot p - f_0(x,\omega) - y \cdot H(x,\omega) \mid x \in \operatorname{dom} H(\cdot,\omega)\} \end{aligned}$$

for $y \in K^\circ$ and $f^*(p,y,\omega) = +\infty$ for $y \notin K^\circ$. If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$, Corollary 5.6 says that the dual problem can be written as

$$\begin{aligned} &\text{maximize} \quad E[\inf_{x\in\mathbb{R}^n}\{f_0(x) + y \cdot H(x) - x \cdot p\}] \quad \text{over } (p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y} \\ &\text{subject to} \qquad y \in K^\circ \quad a.s. \end{aligned} \tag{7.1}$$

To get more explicit expressions for f^* and the dual problem, additional structure is needed; see e.g Example 7.2 below. Theorem 6.3 gives the following.

Theorem 7.1. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (**??**) *and* (7.1) *are feasible, then the following are equivalent:*

1. *x solves* (**??**), *(p,y) solves* (7.1) *and* $\inf$ (**??**) $= \sup$ (7.1)*;*
2. *x is feasible in* (**??**), *(p,y) is feasible in* (7.1) *and*

$$\begin{gathered} p \in \partial_x[f_0 + y \cdot H](x), \\ H(x) \in K, \quad y \in K^\circ, \quad y \cdot H(x) = 0 \end{gathered}$$

almost surely.

Proof. By Theorem 6.3, it suffices to note that, when $(x,y) \in \operatorname{dom} l$, we have

$$0 \in \partial_y[-l](x,y) = -H(x) + N_{K^\circ}(y),$$

if and only if $H(x) \in \partial\delta_{K^\circ}(y)$. By ?THM? **??**, this is equivalent to the given complementarity condition. □

The more general composite format in Example 3.47 can be treated in an analogous way. In case of linear stochastic programming, the dual can be written down explicitly in terms of the problem data.

Example 7.2 (Linear stochastic programming)**.** *Consider the problem*

$$\begin{array}{ll} \text{minimize} & E[x \cdot c] \quad \text{over}\, x \in \mathcal{N} \\ \text{subject to} & Ax + b \in K \quad a.s. \end{array} \tag{7.2}$$

from Example 3.48 *and assume that there exists* $(x,u) \in \mathcal{X}\times\mathcal{U}$ *such that* $E[x\cdot c] < \infty$ *and* $Ax + u + b \in K$ *almost surely. The dual problem becomes*

$$\begin{array}{ll} \text{maximize} & E[b \cdot y] \quad \text{over}\;\; p \in \mathcal{X}_a^\perp,\;\; y \in \mathcal{Y}, \\ \text{subject to} & \quad A^*y + c = p,\;\; y \in K^\circ \quad a.s. \end{array}$$

and the scenariowise KKTR-conditions

$$\begin{gathered} A^*y + c = p, \\ Ax + b \in K, \quad y \in K^\circ, \quad (Ax+b)\cdot y = 0, \end{gathered}$$

where A^* *is the scenariowise transpose of* A*.*

Proof. This is a special case of (**??**) with $f_0(x,\omega) = c(\omega)\cdot x$ and $f_j(x,\omega) = a_j(\omega)\cdot x + b_j(\omega)$ for $j = 1,\dots,m$. We get

$$l(x,y,\omega) = x\cdot c(\omega) + y \cdot A(\omega)x + y\cdot b(\omega) - \delta_{K^\circ}(y)$$

and

$$\begin{aligned} f^*(p,y,\omega) &= \sup_{x\in\mathbb{R}^n}\{x\cdot p - l(x,y,\omega)\} \\ &= \begin{cases} -y\cdot b(\omega) & \text{if } y \in K^\circ \text{ and } A^*(\omega)y + c(\omega) = p, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

This gives the dual problem while the KKTR-conditions follow directly from Theorem 7.1. □

The stochastic linear programming problem satisfies the assumptions of Theorem 6.7 under natural conditions. We will denote the *adapted projection* of an integrable process $u = (u_t)_{t=0}^T$ by

$$^a u := (E_t u_t)_{t=0}^T.$$

Example 7.3 (Linear stochastic programming, reduced dual)**.** *In the setting of Example* 7.2 *assume that* $c \in \mathcal{V}$ *and* $A^*y \in \mathcal{V}$ *for all* $y \in \mathcal{Y}$*. Then, the optimum value of the dual problem equals that of the* reduced dual problem

$$\begin{array}{lll} \text{maximize} & E[b \cdot y] \quad \text{over } y \in \mathcal{Y}, \\ \text{subject to} & {}^a(A^*y + c) = 0, \; y \in K^\circ \quad a.s. \end{array} \tag{7.3}$$

and a pair (p, y) *solves the dual if and only if* y *solves* (7.3) *and*

$$p = A^*y + c - {}^a(A^*y + c).$$

If (7.2) *and* (7.3) *are feasible, then the following are equivalent:*

1. x *solves* (7.2)*,* y *solves* (7.3) *and* $\inf$ (**??**) $= \sup$ (6.3)*;*
2. x *is feasible in* (7.2)*,* y *is feasible in* (7.3) *and*

$$\begin{gathered} {}^a(A^*y + c) = 0, \\ Ax + b \in K, \quad y \in K^\circ, \quad (Ax + b) \cdot y = 0. \end{gathered}$$

Proof. This fits the format of Theorem 6.7 with $\pi(y) := A^*y + c - {}^a(A^*y + c)$ and $\gamma(y) = y$. Indeed, using the expression for f^* in the proof of Example 7.2, gives

$$\varphi^*(p, y) = \begin{cases} E[-b \cdot y] & \text{if } p \in \mathcal{X}_a^\perp \text{ and } y \in K^\circ \text{ and } A^*y + c = p, \\ +\infty & \text{otherwise.} \end{cases}$$

Since any $p \in \mathcal{X}_a^\perp$ has ${}^a p = 0$, it is clear that Assumption 5.12 is satisfied so the claims follow from Theorem 6.7. □

Remark 7.4. *If, in Example* 7.3*, the elements of* c_t *and the columns* A_t *of* A *corresponding to* x_t *are* $\mathcal{F}_t$*-measurable, then by ?THM?* **??***, the reduced dual can be written as*

$$\begin{array}{ll} \text{maximize} & E[b \cdot y] \quad \text{over } y \in \mathcal{Y}, \\ \text{subject to} & c_t + A_t^* \cdot E_t y = 0 \;\; \forall t, \; y \in K^\circ \quad a.s. \end{array}$$

If, in addition, $b \in \mathcal{U}$ *and* b *is* $\mathcal{F}_T$*-measurable and the space* $\mathcal{Y}$ *is such that* $E_t y \in \mathcal{Y}$ *for every* $y \in \mathcal{Y}$*, then by ?THM?* **??***, we can write this as*

$$\begin{array}{ll} \text{maximize} & E[b \cdot y_T] \quad \text{over } (y_t)_{t=0}^T \in \mathcal{M}^{\mathcal{Y}}, \\ \text{subject to} & c_t + A_t^* \cdot y_t = 0, \; y_t \in K^\circ \;\; \forall t \quad a.s., \end{array}$$

where $\mathcal{M}^{\mathcal{Y}}$ *is the linear space of martingales* $(y_t)_{t=0}^T$ *with* $y_t \in \mathcal{Y}$ *for all* t*.*

7.2 Optimal stopping

Consider again the relaxed optimal stopping problem

$$\begin{array}{ll} \text{maximize} & E\sum_{t=0}^T R_t x_t \quad \text{over}\, x \in \mathcal{N}, \\ \text{subject to} & x \geq 0,\ \sum_{t=0}^T x_t \leq 1 \quad a.s. \end{array}$$

from Sections **??** and **??**. This fits the general duality framework with $n_t = 1$, $m = 1$,

$$f(x,u,\omega) = \begin{cases} -\sum_{t=0}^T x_t R_t(\omega) & \text{if } x \geq 0 \text{ and } \sum_{t=0}^T x_t + u \leq 0, \\ +\infty & \text{otherwise} \end{cases}$$

and $\bar{u} = -1$. Recall that we have flipped signs here to make the maximization problem (**??**) fit the general format (**??**). We get

$$\begin{aligned} l(x,y,\omega) &= \inf_{u \in \mathbb{R}^n} \{f(x,u,\omega) - uy\} \\ &= \inf_{u \in \mathbb{R}^n} \{-\sum_{t=0}^T x_t R_t(\omega) - uy \mid x \geq 0,\ \sum_{t=0}^T x_t + u \leq 0\} \\ &= \begin{cases} -\sum_{t=0}^T x_t R_t(\omega) + y\sum_{t=0}^T x_t + \delta_{\mathbb{R}^n_+}(x) & \text{if } y \geq 0, \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} \sum_{t=0}^T x_t[y - R_t(\omega)] + \delta_{\mathbb{R}^n_+}(x) & \text{if } y \geq 0, \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

and

$$\begin{aligned} f^*(p,y,\omega) &= \sup_{x \in \mathbb{R}^n} \{x \cdot p - l(x,y,\omega)\} \\ &= \sup_{x \in \mathbb{R}^n_+} \sum_{t=0}^T x_t[p_t - y + R_t(\omega)] \\ &= \begin{cases} 0 & \text{if } y \geq 0 \text{ and } p_t + R_t(\omega) \leq y,\ t = 0,\dots,T, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Since $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$, Corollary 5.6 says that the dual of (**??**) can be written as

$$\begin{array}{lll} \text{minimize} & Ey & \text{over}(p,y) \in \mathcal{X}_a^\perp \times \mathcal{Y}_+ \\ \text{subject to} & p_t + R_t \leq y & t = 0,\dots,T \ a.s. \end{array} \tag{7.4}$$

Again, we have changed the sign to conform to the tradition of writing the optimal stopping problem as a maximization problem. It is clear that (**??**) is feasible, and (7.4) is feasible as soon as the pathwise maximum $\max_t R_t$ of R belongs $\mathcal{Y}$. Theorem 6.3 thus gives the following.

Theorem 7.5. *Assume that* $\max_t R_t \in \mathcal{Y}$*. The following are equivalent:*

1. x *solves* (**??**)*,* (p, y) *solves* (7.4) *and there is no duality gap;*
2. $x \in \mathcal{N}$ *and* $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ *and*

$$x_t \geq 0,\ p_t + R_t \leq y,\ x_t(p_t + R_t - y) = 0 \quad t = 0, \dots, T,$$
$$y \geq 0,\ \sum_{t=0}^T x_t \leq 1,\ y(\sum_{t=0}^T x_t - 1) = 0$$

almost surely.

In particular, a stopping time $\tau \in \mathcal{T}$ *solves* (**??**) *and* $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}_+$ *solves the* (7.4) *if and only if* $p_t + R_t \leq y$ *for all* t *and* $p_\tau + R_\tau = y$ *almost surely. Here,* $p_{T+1} := 0$*.*

Proof. The scenariowise KKTR-condition in Theorem 6.3 can be written as

$$p_t + R_t - y \in N_{\mathbb{R}_+}(x_t) \quad t = 0, \dots, T,$$
$$\sum_{t=0}^T x_t - 1 \in N_{\mathbb{R}_+}(y).$$

This is equivalent to the conditions given in the statement; see ?THM? **??**. The second claim thus follows from Theorem 5.7 and Corollary 5.10. The last claim follows from the fact that a $\tau \in \mathcal{T}$ solves the optimal stopping problem (**??**) if and only if the process $x \in \mathcal{N}$ given by

$$x_t = \begin{cases} 1 & \text{if } t = \tau, \\ 0 & \text{if } t \neq \tau \end{cases}$$

is optimal in (**??**); see the beginning of Section **??**. □

Under mild conditions, the assumptions of Theorem 6.7 on the reduced dual problem hold with the mappings $\pi : \mathcal{Y} \to \mathcal{V}$ and $\gamma : \mathcal{Y} \to \mathcal{Y}$ given by $\pi(y) = (y - E_t y)_{t=0}^T$ and $\gamma(y) = E_T y$, respectively. Combining Theorem 7.5 with Theorem 6.7 thus gives the following.

Corollary 7.6 (Reduced dual)**.** *Assume that* $R_t \in \mathcal{Y}$ *and* $E_t\mathcal{Y} \subseteq \mathcal{Y} \subseteq \mathcal{V}_t$ *for all* t*. The optimum value of* (7.4) *equals that of*

$$\begin{aligned} &\text{minimize} \quad Ey \quad \text{over}\, y \in \mathcal{Y}_+(\mathcal{F}_T) \\ &\text{subject to} \quad R_t \leq E_t y \quad t = 0, \dots, T \ a.s. \end{aligned} \tag{7.5}$$

If (p, y) *solves* (7.4)*, then* $E_T y$ *solves* (7.5)*. If* y *solves* (7.5)*, then* $((y - E_t y)_{t=0}^T, y)$ *solves* (7.4)*. An* $x \in \mathcal{N}$ *solves* (**??**)*,* $y \in \mathcal{Y}_+(\mathcal{F}_T)$ *solves* (7.5) *and* *there is no duality gap if and only if*

$$x_t \geq 0,\ R_t \leq E_t y,\ x_t(R_t - E_t y) = 0 \quad t = 0, \dots, T,$$
$$y \geq 0,\ \sum_{t=0}^{T} x_t \leq 1,\ y(\sum_{t=0}^{T} x_t - 1) = 0.$$

Remark 7.7. *The reduced dual in Corollary* 7.6 *can be written as*

$$\begin{array}{ll} \text{minimize} & Ey_0 \quad \text{over}\, y \in \mathcal{M}_+^{\mathcal{Y}} \\ \text{subject to} & R_t \leq y_t \quad t = 0, \dots, T \text{ a.s.}, \end{array} \tag{7.6}$$

where $\mathcal{M}_+^{\mathcal{Y}}$ is the cone of nonnegative martingales y with $y_t \in \mathcal{Y}$ for all $t = 0, \dots, T$. Thus, $x \in \mathcal{N}$ solves the primal, $y \in \mathcal{M}_+^{\mathcal{Y}}$ solves (7.6) *and there is no duality gap if and only if*

$$x_t \geq 0,\ R_t \leq y_t,\ x_t(R_t - y_t) = 0 \quad t = 0, \dots, T,$$
$$y_T \geq 0,\ \sum_{t=0}^{T} x_t \leq 1,\ y_T(\sum_{t=0}^{T} x_t - 1) = 0.$$

In particular, a stopping time $\tau \in \mathcal{T}$ is optimal in (**??**) *and $y \in \mathcal{M}_+^{\mathcal{Y}}$ solves* (7.6) *if and only if $R_t \leq y_t$ for all t and $R_\tau = y_\tau$, where $y_{T+1} := y_T$.*

Chapters **??** and **??** below give sufficient conditions for the absence of a duality gap and the existence of dual solutions, respectively, in the general formulation of (**??**). In optimal stopping, absence of a duality gap and the existence of dual solutions can be proved directly using ?THM? **??**. The argument is based on the Doob decomposition of the Snell envelope of the reward process R. A stochastic process A is said to be *predictable* if A_t is $\mathcal{F}_{t-1}$-measurable for all t.

Lemma 7.8 (Doob decomposition)**.** *Assume that $E_t\mathcal{Y} \subset \mathcal{Y}$ for all t. Given an adapted process $y = (y_t)_{t=0}^T$ with $y_t \in \mathcal{Y}$ for all t, there exist unique processes M and A such that*

$$y = M + A,$$

M is a martingale, A is predictable $A_0 = 0$ and $M_t, A_t \in \mathcal{Y}$ for all t. If y is a supermartingale, A is nonincreasing.

Proof. It suffices to define M and A recursively by $A_0 = 0$, $\Delta A_t = E_{t-1}\Delta y_t$ and $M_0 = y_0$, $\Delta M_t = \Delta y_t - \Delta A_t$. The uniqueness follows from the fact that a process that is both predictable and a martingale is necessarily a constant process. □

Recall from Section **??** that the Snell envelope S of R is the smallest supermartingale that dominates R.

Corollary 7.9 (Snell envelope). *Assume that $R_t \in \mathcal{Y}$ and $E_t\mathcal{Y} \subseteq \mathcal{Y} \subseteq \mathcal{V}_t$ for all t. Then there is no duality gap and the martingale part M of the Snell envelope of R solves* (7.6)*. The optimal stopping times τ are characterised by the condition $R_\tau = M_\tau$, where $M_{T+1} := M_T$.*

Proof. Let S be the Snell envelope of the reward process R. By decomposability of $\mathcal{Y}$, $S_t \in \mathcal{Y}$ and, since S is a supermartingale, it admits the decomposition $S = M + A$ from Lemma 7.8. Since A is nonincreasing and $A_0 = 0$, the martingale M dominates R and $ES_0 = EM_0$ while, by ?THM? **??**, ES_0 equals the optimum value of (**??**). Thus, M solves (7.6) and there is no duality gap. The last claim now follows from that of Remark 7.7. □

Remark 7.10 (Davis–Karatzas duality). *Trivially, the normal integrand*

$$\tilde{f}(x,u,\omega) = \begin{cases} -\sum_{t=0}^T x_t R_t(\omega) & \text{if } x \geq 0,\ \sum_{t=0}^T x_t + u \leq 0 \text{ and } \sum_{t=0}^T x_t \leq 1, \\ +\infty & \text{otherwise} \end{cases}$$

gives rise to the same primal problem as f introduced above. The corresponding Lagrangian integrand becomes

$$\tilde{l}(x,y,\omega) = \begin{cases} \sum_{t=0}^T x_t(y - R_t(\omega)) + \delta_{\mathbb{R}^n_+}(x) + \delta_{\mathbb{R}_-}(\sum_{t=0}^T x_t - 1) & \text{if } y \geq 0, \\ -\infty & \text{otherwise} \end{cases}$$

and the conjugate integrand

$$\begin{aligned} \tilde{f}^*(p,y,\omega) &= \sup_x \{x \cdot p - \tilde{l}(x,y,\omega)\} \\ &= \sup_{x \in \mathbb{R}^n_+} \left\{ \sum_{t=0}^T x_t(p_t - y + R_t(\omega)) \;\middle|\; \sum_{t=0}^T x_t \leq 1 \right\} \\ &= \sup_{t=0,\dots,T} \{p_t - y + R_t(\omega)\}. \end{aligned}$$

Much like in Remark 7.7*, we find the reduced dual problem*

$$\text{minimize} \quad E\left[\sup_{t=0,\dots,T}\{R_t + y_T - y_t\}\right] \quad \text{over } y \in \mathcal{M}^{\mathcal{Y}}_+. \tag{7.7}$$

Note that, unlike f^*, the normal integrand $\tilde{f}^*$ is everywhere finite and it is dominated by f^*. It follows that the optimum value of the dual problem associated with $\tilde{f}$ lies between the original primal and dual optimum values. In particular, if $\inf$ (**??**) $= \sup$ (7.4)*, then the optimum value of* (7.7) *equals that of* (7.4)*. The finiteness of $\tilde{f}^*$ may make* (7.7) *easier to solve numerically. Being sandwiched between the primal and the original dual, its values provide tighter upper bounds for the optimum value of the primal problem.*

7.3 Optimal control

Consider the optimal control problem

$$
\begin{aligned}
&\text{minimize} \quad && E\left[\sum_{t=0}^{T} L_t(X_t, U_t)\right] \quad \text{over } (X,U) \in \mathcal{N}, \\
&\text{subject to} \quad && \Delta X_t = A_t X_{t-1} + B_t U_{t-1} + W_t \quad t = 1, \dots, T \ a.s.
\end{aligned}
\tag{7.8}
$$

and recall that X_t takes values in $\mathbb{R}^N$ and U_t in $\mathbb{R}^M$. Problem (7.8) fits the general duality framework with $x = (X, U)$, $\bar{u} = (W_t)_{t=1}^T$, $n_t = N + M$, $m = T$ and

$$
f(x, u, \omega) = \sum_{t=0}^{T} L_t(X_t, U_t, \omega) + \sum_{t=1}^{T} \delta_{\{0\}}(\Delta X_t - A_t(\omega) X_{t-1} - B_t(\omega) U_{t-1} - u_t).
$$

We thus assume that $\mathcal{X}$ and $\mathcal{U}$ are spaces of $\mathbb{R}^{(T+1)(N+M)}$- and $\mathbb{R}^{TM}$-valued random variables, respectively, and that $(W_1, \dots, W_T) \in \mathcal{U}$. By solidity,

$$
\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_T, \quad \mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T,
$$

where $\mathcal{U}_t$ and $\mathcal{Y}_t$ are solid decomposable spaces of $\mathbb{R}^M$-valued random variables in separating duality under the bilinear form $(u_t, y_t) \mapsto E[u_t \cdot y_t]$. It follows that

$$
\langle u, y \rangle = \sum_{t=1}^{T} E[u_t \cdot y_t].
$$

For simplicity, we assume further that, for all t,

$$
\begin{aligned}
\mathcal{X}_t &= \mathcal{S} \times \mathcal{C}, \qquad & \mathcal{U}_t &= \mathcal{S}, \\
\mathcal{V}_t &= \mathcal{S}' \times \mathcal{C}', \qquad & \mathcal{Y}_t &= \mathcal{S}',
\end{aligned}
$$

where $\mathcal{S}$ and $\mathcal{C}$ are solid decomposable spaces in separating duality with solid decomposable spaces $\mathcal{S}'$ and $\mathcal{C}'$, respectively.

The Lagrangian integrand becomes

$$
\begin{aligned}
l(x, y, \omega) &= \inf_{u \in \mathbb{R}^m} \{f(x, u, \omega) - u \cdot y\} \\
&= \sum_{t=0}^{T} L_t(X_t, U_t, \omega) - \sum_{t=1}^{T} (\Delta X_t - A_t(\omega) X_{t-1} - B_t(\omega) U_{t-1}) \cdot y_t.
\end{aligned}
$$

Using the *integration by parts* formula,

$$
\sum_{t=1}^{T} \Delta X_t \cdot y_t = -\sum_{t=0}^{T} X_t \cdot \Delta y_{t+1},
$$

where $y_0 := y_{T+1} := 0$, we get

$$
\begin{aligned}
-\sum_{t=1}^{T}(\Delta X_t - A_t(\omega)X_{t-1} - B_t(\omega)U_{t-1}) \cdot y_t \\
= \sum_{t=0}^{T} X_t \cdot (\Delta y_{t+1} + A^*_{t+1}(\omega)y_{t+1}) + U_t \cdot B^*_{t+1}(\omega)y_{t+1}), \quad (7.9)
\end{aligned}
$$

where $A_{T+1} := 0$ and $B_{T+1} := 0$. Thus, the Lagrangian integrand can be written as

$$
\begin{aligned}
l(x, y, \omega) = \sum_{t=0}^{T}[L_t(X_t, U_t, \omega) \\
+ (X_t, U_t) \cdot (\Delta y_{t+1} + A^*_{t+1}(\omega)y_{t+1}, B^*_{t+1}(\omega)y_{t+1})]
\end{aligned}
$$

and the conjugate of f becomes

$$
\begin{aligned}
f^*(v, y, \omega) &= \sup_{x \in \mathbb{R}^n} \{x \cdot v - l(x, y, \omega)\} \\
&= \sum_{t=0}^{T} L^*_t(v_t - (\Delta y_{t+1} + A^*_{t+1}(\omega)y_{t+1}, B^*_{t+1}(\omega)y_{t+1}), \omega).
\end{aligned}
$$

As soon as $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$, Theorem 5.5 says that the conjugate of the optimum value function can be written as

$$
\varphi^*(p, y) = E\left[\sum_{t=0}^{T} L^*_t(p_t - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}))\right] + \delta_{\mathcal{X}_a^\perp}(p) \quad (7.10)
$$

and, by Corollary 5.6, the dual problem becomes

$$
\begin{aligned}
&\text{maximize} \quad E\left[\sum_{t=1}^{T} W_t \cdot y_t - \sum_{t=0}^{T} L^*_t(p_t - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}))\right] \\
&\text{over} \qquad (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}.
\end{aligned}
\quad (7.11)
$$

Theorem 6.3 gives the following.

Theorem 7.11. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (7.8) *and* (7.11) *are feasible, then the following are equivalent:*

1. (X, U) *solves* (7.8), (p, y) *solves* (7.11) *and there is no duality gap;*
2. (X, U) *is feasible in* (7.8), (p, y) *is feasible in* (7.11) *and*

$$
p_t - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}) \in \partial L_t(X_t, U_t),
$$

for all t *almost surely.*

The optimality conditions in Theorem 7.11 yield a characterization of the optimal control U as a pointwise minimizer of a "Hamiltonian" function associated with problem (7.8).

Remark 7.12 (Maximum principle)**.** *The scenariowise KKTR-conditions in Theorem* 7.11 *mean that* (X, U) *satisfies the system equations and that*

$$-(\Delta y_{t+1}, 0) \in \partial_{(X_t,U_t)} H_t(X_t, U_t, y_{t+1}) - p_t,$$

where

$$H_t(X_t, U_t, y_{t+1}) := L_t(X_t, U_t) + y_{t+1} \cdot (A_{t+1} X_t + B_{t+1} U_t).$$

This can be written equivalently as

$$\begin{aligned} U_t &\in \operatorname*{argmin}_{U_t \in \mathbb{R}^M} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\}, \\ -\Delta y_{t+1} &\in \partial_{X_t} \bar{H}_t(X_t, p_t, y_{t+1}), \end{aligned}$$

where

$$\bar{H}_t(X_t, p_t, y_{t+1}) := \inf_{U_t \in \mathbb{R}^M} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\}.$$

If, for all $(X_t, U_t, y_{t+1}) \in \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^N$,

$$\partial_{(X_t,U_t)} H_t(X_t, U_t, y_{t+1}) = \partial_{X_t} H_t(X_t, U_t, y_{t+1}) \times \partial_{U_t} H_t(X_t, U_t, y_{t+1}), \quad (7.12)$$

this can be written as

$$\begin{aligned} U_t &\in \operatorname*{argmin}_{U_t \in \mathbb{R}^M} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\}, \\ -\Delta y_{t+1} &\in \partial_{X_t} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\} \end{aligned}$$

almost surely. Condition (7.12) *holds, by ?THM?* **??***, in particular, if* L_t *is of the form*

$$L_t(X, U) = L_t^0(X, U) + L_t^1(X) + L_t^2(U),$$

where all the functions are convex and L_t^0 *is finite and differentiable.*

Proof. The optimality conditions in Theorem 7.11 mean that

$$-(\Delta y_{t+1}, 0) \in \partial f_t(X_t, U_t), \quad (7.13)$$

where $f_t(X_t, U_t) := H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t$. The first claim thus follows from Theorem 2.33 with $v = 0$, $x = U_t$, $u = X_t$ and $F = f_t$. Under (7.12), condition (7.13) can be written as

$$\begin{aligned} -\Delta y_{t+1} &\in \partial_{X_t} f_t(X_t, U_t), \\ 0 &\in \partial_{U_t} f_t(X_t, U_t), \end{aligned}$$

which is the second condition. □

Remark 7.19 below gives a version of the maximum principle which does not involve the shadow price of information p. This will require some extra assumptions on the problem data. Recall that the Köthe dual of a space $\mathcal{U}$ of random variables is the linear space

$$\{y \in L^0 \mid u \cdot y \in L^1 \ \forall u \in \mathcal{U}\}.$$

Assumption 7.13. *The spaces $\mathcal{S}'$ and $\mathcal{C}'$ are the Köthe duals of $\mathcal{S}$ and $\mathcal{C}$, respectively, and, for all t,*

1. *$E_t\mathcal{S} \subseteq \mathcal{S}$ and $E_t\mathcal{C} \subseteq \mathcal{C}$,*
2. *$A_t\mathcal{S} \subseteq \mathcal{S}$ and $B_t\mathcal{C} \subseteq \mathcal{S}$.*

Except for condition 2, Assumption 7.13 holds automatically e.g. in Lebesgue and Orlicz spaces; see the examples in Section 4.1. Condition 2 imposes natural integrability conditions on the matrices. It holds e.g. in spaces of finite moments if the elements of A_t and B_t have finite moments; see ?THM? **??**. By Lemma 4.6 and Lemma 4.5, Assumption 7.13 implies that, for all t,

1. $E_t\mathcal{S}' \subseteq \mathcal{S}'$ and $E_t\mathcal{C}' \subseteq \mathcal{C}'$,
2. $A_t^*\mathcal{S}' \subseteq \mathcal{S}'$ and $B_t^*\mathcal{S}' \subseteq \mathcal{C}'$.

Note that, under Assumption 7.13, $\operatorname{dom} Ef \cap (\mathcal{X}\times\mathcal{U}) \neq \emptyset$ means that $\operatorname{dom} E[\sum_{t=0}^T L_t] \cap \mathcal{X} \neq \emptyset$ for every t. This holds in particular, if each EL_t is proper on $\mathcal{S}\times\mathcal{C}$.

Under Assumption 7.13, the dual problem (7.11) can be simplified using the general techniques of Section 5.2. We will find that problem (7.8) satisfies Assumption 5.12 with different choices for the mapping Π resulting in different restrictions in the dual problem.

Remark 7.14. *Let $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$. Since $x = (X, U)$, the shadow price of information $p \in \mathcal{X}_a^\perp$ can be decomposed as $p = (V, Y)$, where $V \in (\mathcal{S}')^{T+1}$ and $Y \in (\mathcal{C}')^{T+1}$. Under Assumption 7.13, there exists $(\tilde{p}, \tilde{y}) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ that achieves the same dual objective value as (p, y) but the component of $\tilde{p}_t$ corresponding to the state X_t is zero for $t \geq 1$. This is quite natural given that the information constraint on the state X_t for $t \geq 1$ is redundant as observed in ?THM?* **??**.

Proof. Define $\Pi^* : \mathcal{X}\times\mathcal{U} \to \mathcal{X}\times\mathcal{U}$ by $\Pi^*(Z, R, u) = (\tilde{Z}, R, u)$, where $\tilde{Z}$ is given by

$$\begin{aligned}\tilde{Z}_0 &= Z_0,\\ \Delta\tilde{Z}_t &= A_t\tilde{Z}_{t-1} + B_tR_{t-1} + u_t.\end{aligned}$$

Given $(x, z, u) \in \operatorname{dom} F$ and $(\tilde{x}, \tilde{z}, \tilde{u}) = \Pi^*(x, z, u)$, we have

$$\Delta X_t = A_tX_{t-1} + B_tU_{t-1} + u_t$$

so

$$\Delta(X_t - \tilde{Z}_t) = A_t(X_{t-1} - \tilde{Z}_{t-1}) + B_t(U_{t-1} - R_{t-1}).$$

It follows that $x - \tilde{z} \in \mathcal{N}$ so $F(\tilde{x}, \tilde{z}, \tilde{u}) = F(x, z, u)$. We thus have

$$F(x, \Pi^*(z, u)) \leq F(x, z, u) \quad (x, z, u) \in L^0 \times \mathcal{X} \times \mathcal{U}.$$

Recursive application of Lemma 4.5 shows that, under Assumption 7.13, Π^* is continuous so it has a continuous adjoint Π. Clearly, $\Pi^*(0, \bar{u}) \in \mathcal{X}_a \times \{\bar{u}\}$, so Assumption 5.12 holds, by Lemma 6.9. Thus, as observed at the beginning of 79 5.2, we may restrict the dual problem to $\operatorname{rge}\Pi$. Since

$$\ker \Pi^* = \{(Z, R, u) \mid Z_0 = 0,\ R = 0,\ u = 0\},$$

we have

$$\operatorname{cl}\operatorname{rge}\Pi = (\ker \Pi^*)^\perp = \{(V, Y, y) \mid V_t = 0 \quad t = 1, \dots, T\}$$

which completes the proof. □

Recall that the adapted projection ${}^a y$ of a process $y \in \mathcal{Y}$ is defined by ${}^a y_t := E_t y_t$. We will denote the set of adapted processes in $\mathcal{Y}$ by $\mathcal{Y}_a$.

Remark 7.15. *Let Assumption* 7.13 *hold and* $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$. *There exists* $(\tilde{p}, \tilde{y}) \in \mathcal{X}_a^\perp \times \mathcal{Y}_a$ *that achieves the same dual objective value as* (p, y). *In particular, if dual solutions exist, then there exists one where* y *is adapted. If, in addition, each* L_t *is* $\mathcal{F}_{t+1}$*-measurable and each* EL_t *is closed and proper on* $\mathcal{S} \times \mathcal{C}$*, then there exists* $(\tilde{p}, \tilde{y}) \in \mathcal{X}_a^\perp \times \mathcal{Y}_a$ *such that* $\tilde{p}_t$ *is* $\mathcal{F}_{t+1}$*-measurable and* $(\tilde{p}, \tilde{y})$ *achieves dual objective value at least as good as* (p, y).

Proof. Let $\tilde{y} = {}^a y$ and

$$\tilde{p}_t = p_t + (\Delta\tilde{y}_{t+1} + A^*_{t+1}\tilde{y}_{t+1}, B^*_{t+1}\tilde{y}_{t+1}) - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}).$$

It is clear that the dual objective values are the same at (p, y) and $(\tilde{p}, \tilde{y})$, while Lemma 4.7 implies $\tilde{p} \in \mathcal{X}_a^\perp$. Under the additional assumptions, for any $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}_a$, Theorem 4.13 implies

$$\begin{aligned} &E\left[\sum_{t=0}^T L_t^*(p_t - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}))\right] \\ &\geq E\left[\sum_{t=0}^T L_t^*(E_{t+1}[p_t] - (\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}))\right], \end{aligned}$$

which completes the proof. □

Theorem 6.7 yields the following.

Corollary 7.16 (Reduced dual)**.** *Assume that each L_t is $\mathcal{F}_t$-measurable, each EL_t is closed and proper on $\mathcal{S}\times\mathcal{C}$ and that Assumption 7.13 holds. Then the optimum value of the dual problem (7.11) equals that of the* reduced dual problem

$$\underset{y\in\mathcal{Y}_a}{\text{maximize}}\ E\left[\sum_{t=1}^T W_t\cdot y_t - \sum_{t=0}^T L_t^*(-E_t(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}))\right] \quad (7.14)$$

and the dual has a solution if and only if the reduced dual has a solution. If $\operatorname{dom} Ef\cap(\mathcal{X}\times\mathcal{U})\neq\emptyset$ *and (7.8) and (7.14) are feasible, then the following are equivalent:*

1. *(X,U) solves (7.8), y solves (7.14) and* $\inf(7.8)=\sup(7.14)$*;*
2. *(X,U) is feasible in (7.8), y is feasible in (7.14) and*

$$-E_t(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1})\in\partial L_t(X_t,U_t)$$

for all t almost surely.

Proof. By Assumption 7.13, the mappings $\pi:\mathcal{Y}\to\mathcal{V}$ and $\gamma:\mathcal{Y}\to\mathcal{Y}$ are well-defined by $\gamma(y)={}^a y$ and

$$\pi(y)_t=(\Delta\,{}^a y_{t+1}+A_{t+1}^*\,{}^a y_{t+1},B_{t+1}^*\,{}^a y_{t+1})-E_t(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}).$$

It suffices to show that they satisfy the assumptions of Theorem 6.7. Applying the Jensen's inequality in Theorem 4.13 to the expression of φ^* in (7.10) gives, for every $(p,y)\in\mathcal{X}_a^\perp\times\mathcal{Y}$,

$$\begin{aligned}\varphi^*(p,y)&=E\left[\sum_{t=0}^T L_t^*(p_t-(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}))\right]\\&\geq E\left[\sum_{t=0}^T L_t^*(-E_t[\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}])\right]\\&=\varphi^*(\pi(y),\gamma(y)),\end{aligned}$$

which is the first condition in Assumption 5.12. Since $\bar u=W$ is adapted, the last condition in Assumption 5.12 holds by ?THM? **??**. □

The following gives a more functional analytic proof of Corollary 7.16. The argument is based on establishing the properties of the primal problem in Lemma 5.16.

Remark 7.17. *We saw in the derivation of the Lagrangian integrand that the random linear mapping*

$$K(x,\omega):=(\Delta X_t-A_t(\omega)X_{t-1}-B_t(\omega)U_{t-1})_{t=1}^T$$

has scenariowise adjoint given by

$$K^*(y,\omega) = -(\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1})_{t=0}^T.$$

By Lemma 4.5*, Assumption* 7.13 *implies that the pointwise application of* K *induces a continuous linear mapping* $\mathcal{K} : \mathcal{X} \to \mathcal{U}$ *and that its adjoint* $\mathcal{K}^* : \mathcal{Y} \to \mathcal{V}$ *is given by pointwise application of* K^**.*

Given $(x, z, u) \in \operatorname{dom} F$*, we have* $x - z \in \mathcal{N}$ *and* $\mathcal{K}(x) = u$*. Jensen's inequality gives*

$$EL_t(E_tX_t, E_tU_t) \leq EL_t(X_t, U_t)$$

while

$$\mathcal{K}({}^a x) = \mathcal{K}(x) + \mathcal{K}({}^a x - x) = u + \mathcal{K}({}^a z - z) = u + \mathcal{K}({}^a z) - \mathcal{K}(z).$$

Since $\mathcal{K}({}^a x)$ *and* $\mathcal{K}({}^a z)$ *are adapted, we have*

$$\mathcal{K}({}^a x) = {}^a u + \mathcal{K}({}^a z) - {}^a\mathcal{K}(z).$$

Thus, the conditions of Lemma 5.16 *are satisfied with* $\xi(x, z, u) = {}^a x$ *and*

$$\Pi^*(z, u) = (0, {}^a u + \mathcal{K}({}^a z) - {}^a\mathcal{K}(z)).$$

We have

$$\begin{aligned}\langle \Pi^*(z, u), (p, y)\rangle &= \langle {}^a u + \mathcal{K}({}^a z) - {}^a\mathcal{K}(z), y\rangle \\ &= \langle z, {}^a\mathcal{K}^*(y) - \mathcal{K}^*({}^a y)\rangle + \langle u, {}^a y\rangle\end{aligned}$$

so the adjoint of Π^* *is given by* $\Pi(p, y) = ({}^a\mathcal{K}^*(y) - \mathcal{K}^*({}^a y), {}^a y)$*. This satisfies the assumptions of Theorem* 6.7*.*

Remark 7.18. *Even if the normal integrands* L_t *are not* $\mathcal{F}_t$*-measurable, one can apply Corollary* 7.16 *to the optimal control problem where each* L_t *has been replaced by its conditional expectation* E_tL_t*. The corresponding optimality conditions would then become*

$$-E_t(\Delta y_{t+1} + A^*_{t+1}y_{t+1}, B^*_{t+1}y_{t+1}) \in \partial(E_tL_t)(X_t, U_t)$$

for all t *almost surely. By Fenchel's inequality, dual feasibility implies that* $L_t(X_t, U_t)$ *are quasi-integrable for every* $(X, U) \in \mathcal{X}$ *so, by ?THM?* **??** *and ?THM?* **??***,*

$$E\left[\sum_{t=0}^T L_t(X_t, U_t)\right] = E\left[\sum_{t=0}^T (E_tL_t)(X_t, U_t)\right] \quad \forall (X, U) \in \mathcal{X}_a.$$

The two control problems thus coincide over the space $\mathcal{X}_a$*.*

Remark 7.19 (Maximum principle in reduced form)**.** *The scenariowise optimality condition in Corollary* 7.16 *can be written as*

$$-(E_t\Delta y_{t+1},0)\in\partial_{(X,U)}H_t(X_t,U_t,y_{t+1}),$$

where

$$H_t(X_t,U_t,y_{t+1}):=L_t(X_t,U_t)+E_t[A^*_{t+1}y_{t+1}]\cdot X_t+E_t[B^*_{t+1}y_{t+1}]\cdot U_t.$$

As in Remark 7.12*, the optimality conditions can thus be written as*

$$\begin{aligned}U_t&\in\mathop{\mathrm{argmin}}_{U_t\in\mathbb{R}^M}H_t(X_t,U_t,y_{t+1}),\\ -E_t\Delta y_{t+1}&\in\partial_X\bar H_t(X_t,y_{t+1}),\end{aligned}$$

where

$$\bar H_t(X_t,y_{t+1}):=\inf_{U_t\in\mathbb{R}^M}H_t(X_t,U_t,y_{t+1}).$$

7.4 Problems of Lagrange

Consider again problem (**??**)

$$\text{minimize}\quad E\left[\sum_{t=0}^T K_t(x_t,\Delta x_t)\right]\quad\text{over}\,x\in\mathcal{N}$$

from Section **??**. This fits the general duality framework with $\bar u=0$ and

$$f(x,u,\omega)=\sum_{t=0}^T K_t(x_t,\Delta x_t+u_t,\omega).$$

We thus assume that both $\mathcal{X}$ and $\mathcal{U}$ are spaces of $\mathbb{R}^{(T+1)d}$-valued random variables. For simplicity, we assume that

$$\mathcal{X}_t=\mathcal{S},\quad\mathcal{V}_t=\mathcal{S}',\quad\mathcal{U}=\mathcal{X},\quad\mathcal{Y}=\mathcal{V},$$

where $\mathcal{S}$ and $\mathcal{S}'$ are solid decomposable spaces in separating duality under the bilinear form $\langle x_t,v_t\rangle:=E[x_t\cdot v_t]$.

The Lagrangian integrand becomes

$$\begin{aligned}l(x,y,\omega)&=\inf_{u\in\mathbb{R}^m}\{f(x,u,\omega)-u\cdot y\}\\&=\sum_{t=0}^T(\Delta x_t\cdot y_t+H_t(x_t,y_t,\omega))\\&=\sum_{t=0}^T(-x_t\cdot\Delta y_{t+1}+H_t(x_t,y_t,\omega)),\end{aligned}$$

where we define $y_{T+1} := 0$ and

$$H_t(x_t, y_t, \omega) := \inf_{u_t \in \mathbb{R}^d} \{K_t(x_t, u_t, \omega) - u_t \cdot y_t\}.$$

The function H_t is called the *Hamiltonian* associated with K_t. It has similar properties as the Lagrangian integrand in Section 5.1. In particular, $H_t(x_t, y_t, \omega)$ is convex in x_t and concave in y_t. The conjugate integrand can be written as

$$\begin{aligned} f^*(v, y, \omega) &= \sup_{x \in \mathbb{R}^n} \{x \cdot v - l(x, y, \omega)\} \\ &= \sup_{x \in \mathbb{R}^n} \left\{ \sum_{t=0}^{T} x_t \cdot v_t - \sum_{t=0}^{T} (-x_t \cdot \Delta y_{t+1} + H_t(x_t, y_t, \omega)) \right\} \\ &= \sum_{t=0}^{T} K_t^*(v_t + \Delta y_{t+1}, y_t, \omega). \end{aligned}$$

If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$, Theorem 5.5 says that the conjugate of the optimum value function can be written as

$$\varphi^*(p, y) = E\left[\sum_{t=0}^{T} K_t^*(p_t + \Delta y_{t+1}, y_t)\right] + \delta_{\mathcal{X}_a^\perp}(p) \tag{7.15}$$

and, by Corollary 5.6, the dual problem becomes

$$\text{maximize} \qquad E\left[-\sum_{t=0}^{T} K_t^*(p_t + \Delta y_{t+1}, y_t)\right] \qquad \text{over} (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}. \tag{7.16}$$

Theorem 5.7 and Corollary 5.10 now give the following.

Theorem 7.20. *If* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (**??**) *and* (7.16) *are feasible, then the following are equivalent:*

1. *x solves* (**??**), *(p, y) solves* (7.16) *and there is no duality gap;*
2. *x is feasible in* (**??**), *(p, y) is feasible in* (7.16) *and, for all t,*

   $$(p_t + \Delta y_{t+1}, y_t) \in \partial K_t(x_t, \Delta x_t)$$

   almost surely;
3. *x is feasible in* (**??**), *(p, y) is feasible in* (7.16) *and, for all t,*

   $$\begin{aligned} p_t + \Delta y_{t+1} &\in \partial_x H_t(x_t, y_t), \\ \Delta x_t &\in \partial_y [-H_t](x_t, y_t), \end{aligned}$$

   almost surely.

The subgradient condition in part 2 of Theorem 7.21 is known as *Euler-Lagrange condition* for (**??**) while the conditions in 3 are known as *Hamiltonian conditions.* To clarify the connection with classical formulations, consider the deterministic case where $p = 0$ and assume that both K_t and H_t are differentiable. Condition 3 can then be written as

$$\Delta y_{t+1} = \nabla_x H_t(x_t, y_t), \quad \Delta x_t = -\nabla_y H_t(x_t, y_t)$$

while the conditions in 2 become

$$\Delta y_{t+1} = \nabla_x K_t(x_t, \Delta x_t), \quad y_t = \nabla_u K_t(x_t, \Delta x_t),$$

which is a discrete-time version of the classical Euler-Lagrange equation

$$\nabla_x K_t(x_t, \dot{x}_t) = \frac{d}{dt}\nabla_u K_t(x_t, \dot{x}_t).$$

Much as in Remark 7.12, one could formulate the optimality conditions in part 3 of Theorem 7.21 in the form of a maximum principle, where the optimal solutions x are pointwise minimizers of a linearly tilted Hamiltonian.

Under the following assumption, the optimality conditions can be written without the shadow price of information p, much as in Corollary 7.16.

Assumption 7.21. *The space $\mathcal{S}'$ is the Köthe dual of $\mathcal{S}$ and $E_t\mathcal{S} \subseteq \mathcal{S}$ for all t.*

By Lemma 4.6, Assumption 7.22 implies that $E_t\mathcal{S}' \subseteq \mathcal{S}'$ for all t. As before, we will denote the set of adapted processes in $\mathcal{Y}$ by $\mathcal{Y}_a$. The adapted projection ${}^a y$ of a process $y \in \mathcal{Y}$ is defined by ${}^a y_t = E_t y_t$.

Theorem 6.7 yields the following.

Corollary 7.22 (Reduced dual)**.** *Assume that* $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$*, Assumption* 7.22 *holds, that* (**??**) *and* (7.16) *are feasible and that, for all t, K_t is $\mathcal{F}_t$-measurable and EK_t is proper on $\mathcal{S} \times \mathcal{S}$. Then the optimum value of the dual problem* (7.16) *equals that of the reduced dual problem*

$$\text{maximize} \qquad E[-\sum_{t=0}^{T} K_t^*(E_t\Delta y_{t+1}, y_t)] \quad \text{over } y \in \mathcal{Y}_a$$

and the dual has a solution if and only if the reduced dual has a solution. If the reduced dual has a solution, then an x is optimal if and only if it is feasible and there is a y feasible in the reduced dual such that

$$E_t\Delta y_{t+1} \in \partial_x H_t(x_t, y_t),$$
$$\Delta x_t \in \partial_y[-H_t](x_t, y_t)$$

almost surely.

Proof. Under Assumption 7.22, the mappings $\pi : \mathcal{Y} \to \mathcal{V}$ and $\gamma : \mathcal{Y} \to \mathcal{Y}$ are well-defined by $\gamma(y) = {}^a y$ and

$$\pi(y)_t = E_t[\Delta y_{t+1}] - \Delta\, {}^a y_{t+1}.$$

It suffices to show that they satisfy the assumptions of Theorem 6.7. Applying the Jensen's inequality in Theorem 4.13 to the expression of φ^* in (7.15) gives, for every $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$,

$$\begin{aligned}
\varphi^*(p, y) &= E\left[\sum_{t=0}^{T} K_t^*(p_t + \Delta y_{t+1}, y_t)\right] \\
&\geq E\left[\sum_{t=0}^{T} K_t^*(E_t[\Delta y_{t+1}], E_t[y_t])\right] \\
&= \varphi^*(\pi(y), \gamma(y)),
\end{aligned}$$

which is the first condition in Assumption 5.12. The last condition in Assumption 5.12 holds trivially since $\bar{u} = 0$. □

Remark 7.23. *By Lemma* 7.8*, any* $y \in \mathcal{Y}_a$ *has a unique decomposition* $y = M + A$ *where* M *is a martingale and* A *is predictable with* $A_0 = 0$*. Defining* $A_{T+1} := -M_T$ *the reduced dual problem in Corollary* 7.23 *can thus be written without conditional expectations as*

$$\text{maximize} \quad E[-\sum_{t=0}^{T} K_t^*(\Delta A_{t+1}, M_t + A_t)] \quad \text{over } (M, A) \in \mathcal{M}^{\mathcal{Y}} \times \mathcal{Y}_p,$$

where $\mathcal{M}^{\mathcal{Y}} \subset \mathcal{Y}$ *and* $\mathcal{Y}_p \subset \mathcal{Y}$ *are the linear spaces of martingales and predictable processes, respectively. Accordingly, the optimality conditions in Corollary* 7.23 *can be written as*

$$\begin{aligned}
\Delta A_{t+1} &\in \partial_x H_t(x_t, M_t + A_t), \\
\Delta x_t &\in \partial_y [-H_t](x_t, M_t + A_t).
\end{aligned}$$

Example 7.24 (Optimal stopping)**.** *Let Assumption* 7.22 *hold and let* R *be an adapted process with* $R_t \in \mathcal{S}'$ *for all* t*. The relaxed optimal stopping problem from Section* **??** *can be written as*

$$\underset{x \in \mathcal{N}_+}{\text{maximize}} \quad E\sum_{t=0}^{T} R_t \Delta x_t \quad \text{subject to} \quad \Delta x \geq 0,\ x_T \leq 1 \ a.s.$$

This is a problem of Lagrange with

$$K_t(x_t, u_t) = -R_t u_t + \delta_{\mathbb{R}_-}(x_t - 1) + \delta_{\mathbb{R}_+}(u_t).$$

We get

$$
\begin{aligned}
K_t^*(v_t, y_t) &= \sup_{x_t, u_t \in \mathbb{R}} \{x_t \cdot v_t + u_t \cdot y_t - K_t(x_t, u_t)\} \\
&= \sup_{x_t, u_t \in \mathbb{R}} \{x_t \cdot v_t + u_t \cdot y_t + R_t u_t \mid x_t \leq 1,\ u_t \geq 0\} \\
&= \begin{cases} v_t & \text{if } v_t \geq 0 \text{ and } R_t + y_t \leq 0, \\ +\infty & \text{otherwise,} \end{cases}
\end{aligned}
$$

so the reduced dual in Corollary 7.23 becomes

$$
\begin{aligned}
&\text{maximize} && Ey_0 \ \text{over} y \in \mathcal{Y}_a \\
&\text{subject to} && E_t[\Delta y_{t+1}] \geq 0, \\
& && R_t + y_t \leq 0,
\end{aligned}
$$

where $y_{T+1} := 0$*. With the change of variables* $S = -y$*, this can be written as*

$$
\begin{aligned}
&\text{minimize} && ES_0 \ \text{over} S \in \mathcal{Y}_a \\
&\text{subject to} && E_t[\Delta S_{t+1}] \leq 0, \\
& && R_t \leq S_t.
\end{aligned}
$$

Here one minimizes the expectation of the initial value of a supermartingale that dominates the reward process. In the dual problem in Corollary 7.6*, one minimizes expectation of the initial value of a martingale that dominates the reward process. By Corollary* 7.9*, the martingale part of the Snell envelope* S *of* R *solves the dual in Corollary* 7.6 *while the above dual is solved by the Snell envelope itself. Indeed, if the dual had a solution* S *better than the Snell envelope, then* S *would be a supermartingale that dominates* R *(recall that* $S_{T+1} = y_{T+1} = 0$*) and it would be strictly smaller than the Snell envelope at* $t = 0$*, contradicting the definition of the Snell envelope.*

The following is an analogue of the first part of ?THM? **??**.

Corollary 7.25 (Cost-to-go functions)**.** *Recall the Bellman equations*

$$
\begin{aligned}
\tilde{V}_T &= 0, \\
V_t &= E_t \tilde{V}_t, \\
\tilde{V}_{t-1}(x_{t-1}, \omega) &= \inf_{x_t \in \mathbb{R}^d} \{(E_t K_t)(x_t, \Delta x_t, \omega) + V_t(x_t, \omega)\}
\end{aligned}
$$

from Section **??***. If* x *and* (p, y) *satisfy the optimality conditions in Theorem* 7.21 *and if* $K_t(x_t, \Delta x_t)$ *and* $\tilde{V}_t(x_t)$ *are integrable, then*

$$
-E_t y_{t+1} \in \partial V_t(x_t)
$$

for all t *almost surely.*

Proof. Since $y_{T+1} = 0$, by definition, the claim holds trivially for $t = T$. Assume that $-E_t y_{t+1} \in \partial V_t(x_t)$ and let

$$F_t(x_t, x_{t-1}, \omega) := (E_t K_t)(x_t, \Delta x_t, \omega) + V_t(x_t, \omega).$$

If x and (p, y) satisfy the optimality conditions in Theorem 7.21 and if $K_t(x_t, \Delta x_t)$ is integrable, ?THM? **??** gives

$$(E_t \Delta y_{t+1}, E_t y_t) \in \partial (E_t K_t)(x_t, \Delta x_t).$$

By ?THM? **??**,

$$(0, -E_t y_t) \in \partial F_t(x_t, x_{t-1}).$$

By Theorem 2.33, this implies $-E_t y_t \in \partial \tilde{V}_{t-1}(x_{t-1})$. If $\tilde{V}_{t-1}(x_{t-1})$ is integrable, ?THM? **??** gives $-E_{t-1} y_t \in \partial V_{t-1}(x_{t-1})$. The claim now follows by induction on t. □

Much as in ?THM? **??**, converse of Corollary 7.26 holds under additional conditions.

7.5 Financial mathematics

Consider again problem (3.4)

$$\begin{aligned} &\text{minimize} \quad EV\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}, \\ &\text{subject to} \quad x_t \in D_t \quad t = 0, \dots, T \ a.s. \end{aligned}$$

from Section **??**. This fits the general duality framework with $\bar{u} = c$ and

$$f(x, u, \omega) = V\left(u - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega), \omega\right) + \sum_{t=0}^{T} \delta_{D_t(\omega)}(x_t).$$

We thus assume that $\mathcal{X}_t$ and $\mathcal{U}$ are spaces of $\mathbb{R}^J$- and $\mathbb{R}$-valued random variables, respectively, and that $c \in \mathcal{U}$. For simplicity, we also assume that, for all t,

$$\mathcal{X}_t = \mathcal{S} \quad \text{and} \quad \mathcal{V}_t = \mathcal{S}'$$

where $\mathcal{S}$ and $\mathcal{S}'$ are solid decomposable spaces in separating duality. We continue with the assumptions of Section **??** on the model. In particular, the loss function V is convex, nondecreasing and nonconstant, $0 \in D_t$ almost surely for all t, and $D_T = \{0\}$. For convenience, we will also assume that $V(0) = 0$.

The Lagrangian integrand can be written as

$$\begin{aligned} l(x, y, \omega) &= \inf_{u \in \mathbb{R}} \{f(x, u, \omega) - uy\} \\ &= -y \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega) - V^*(y, \omega) + \sum_{t=0}^{T} \delta_{D_t(\omega)}(x_t). \end{aligned}$$

and, since $D_T = \{0\}$, the conjugate of f becomes

$$\begin{aligned} f^*(v, y, \omega) &= \sup_{x \in \mathbb{R}^J} \{x \cdot v - l(x, y, \omega)\} \\ &= V^*(y, \omega) + \sum_{t=0}^{T-1} \sigma_{D_t(\omega)}(v_t + y \Delta s_{t+1}(\omega)), \end{aligned}$$

where

$$\sigma_{D_t(\omega)}(v) := \sup_{x \in D_t(\omega)} x \cdot v$$

is the support function of $D_t(\omega)$. Since Ef is finite at the origin, Theorem 5.5 says that the conjugate of the optimum value function can be written as

$$\varphi^*(p, y) = E\left[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(p_t + y\Delta s_{t+1})\right] + \delta_{\mathcal{X}_a^\perp}(p) \tag{7.17}$$

and, by Corollary 5.6, the dual problem becomes

$$\text{maximize } E\left[cy - V^*(y) - \sum_{t=0}^{T-1} \sigma_{D_t}(p_t + y\Delta s_{t+1})\right] \text{ over} (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}. \tag{7.18}$$

Theorem 5.7 and Corollary 5.10 now give the following.

Theorem 7.26. *If* (3.4) *and* (7.18) *are feasible, then the following are equivalent:*

1. x *solves* (3.4), (p, y) *solves* (7.18) *and there is no duality gap;*
2. x *is feasible in* (3.4), (p, y) *is feasible in* (7.18) *and*

$$\begin{aligned} y &\in \partial V(u - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}), \\ p_t + y\Delta s_{t+1} &\in N_{D_t}(x_t) \quad t = 0, \dots, T-1 \end{aligned}$$

almost surely.

For the rest of the section, we focus on the reduced dual problem. The following assumption implies the existence of the mappings π and γ in Theorem 6.7.

Assumption 7.27. $\mathcal{Y}$ *is the Köthe dual of* $\mathcal{U}$ *and, for all* t, $\mathcal{X}_t = L^\infty$, $\mathcal{V}_t = L^1$ *and, for every* $t = 0, \dots, T$

1. $E_t \mathcal{U} \subseteq \mathcal{U}$,
2. $\Delta s_{t+1} \in \mathcal{U}$,

where $s_{T+1} := 0$.

Part 1 of Assumption 7.28 holds automatically, e.g., in Lebesgue and Orlicz spaces; see ?THM? **??** in Section **??**. By Lemma 4.6 and Lemma 4.5, Assumption 7.28 implies that, for all t,

1. $E_t\mathcal{Y} \subseteq \mathcal{Y}$,
2. $y\Delta s_{t+1} \in L^1$ for all $y \in \mathcal{Y}$.

Lemma 7.28. *Under Assumption* 7.28*, the mappings* $\pi : \mathcal{Y} \to \mathcal{V}$ *and* $\gamma : \mathcal{Y} \to \mathcal{Y}$ *defined by*

$$\pi(y) := (E_t[y\Delta s_{t+1}] - y\Delta s_{t+1})_{t=0}^T,$$
$$\gamma(y) := y$$

are continuous and satisfy the assumptions of Theorem 6.7*.*

Proof. By Lemma 4.5 and Lemma 4.6, Assumption 7.28 implies that the mapping π is continuous. Applying the Jensen's inequality in Theorem 4.13 to the expression of φ^* in (7.17) gives, for every $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$,

$$\begin{aligned}\varphi^*(p, y) &= E[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(p_t + y\Delta s_{t+1})] \\ &\geq E[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])] \\ &= \varphi^*(\pi(y), \gamma(y))\end{aligned}$$

which is the first condition in Assumption 5.12. The last condition in Assumption 5.12 holds trivially since γ is the identity mapping. □

The following is a direct consequence of Lemma 7.29 and Theorem 6.7.

Theorem 7.29 (Reduced dual)**.** *Let Assumption* 7.28 *hold and* π *be as in Lemma* 7.29*. Then the optimum value of* (7.18) *coincides with that of the problem*

$$\text{maximize} \quad E\left[cy - V^*(y) - \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])\right] \quad \text{over } y \in \mathcal{Y} \qquad (7.19)$$

and if $y \in \mathcal{Y}$ *solves* (7.19)*, then* $(\pi(y), y)$ *solves* (7.18)*. If* $(p, y) \in \mathcal{V} \times \mathcal{Y}$ *solves* (7.18) *then* y *solves* (7.18)*. If* (3.4) *and* (7.19) *are feasible, then the following are equivalent:*

1. *x solves* (3.4)*, y solves* (7.19) *and* $\inf(3.4) = \sup(7.19)$*;*

2. x is feasible in (3.4)*, y is feasible in* (7.19) *and*

$$y \in \partial V(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}),$$
$$E_t[y\Delta s_{t+1}] \in N_{D_t}(x_t) \quad t = 0, \dots, T$$

almost surely.

Since V is nondecreasing, the first subgradient condition in Theorem 7.30 implies that y is almost surely nonnegative. In the absence of portfolio constraints, the last condition reads $E_t[y\Delta s_{t+1}] = 0$. When $y \neq 0$, this is means, by ?THM? **??**, that s is a martingale under the probability measure Q defined by $dQ/dP := y/Ey$.

Combining Theorem 6.7, Theorem 6.8 and Lemma 7.29 gives the following.

Theorem 7.30. *Assume that $\bar{\varphi}$ is closed at $(0, c)$ for all $c \in \mathcal{U}$ and that Assumption* 7.28 *holds. Then*

$$\bar{\varphi}(0, \cdot)^*(y) = E\left[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])\right].$$

Proof. By Lemma 7.29 and Theorem 6.7,

$$g(y) := \inf_{p \in \mathcal{V}} \varphi^*(p, y) = \bar{\varphi}^*(\pi(y), y) = E\left[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])\right],$$

where π is continuous. Thus, g is closed as a composition of a continuous linear mapping with a closed convex function. Thus, the claim follows from Theorem 6.8. □

When $V = \delta_{\mathbb{R}_-}$, the optimum value function of (3.4) becomes $\delta_{\tilde{\mathcal{C}}}$, where

$$\tilde{\mathcal{C}} := \left\{ (z, u) \in \mathcal{X} \times \mathcal{U} \,\middle|\, \exists x \in L^0 : \; x - z \in \mathcal{N}, \; x_t \in D_t, \; \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \geq u \quad a.s. \right\}.$$

The set

$$\begin{aligned} \mathcal{C} :=& \{c \in \mathcal{U} \mid (0, c) \in \tilde{\mathcal{C}}\} \\ =& \left\{ u \in \mathcal{U} \,\middle|\, \exists x \in \mathcal{N}, \; x_t \in D_t, \; \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \geq u \quad a.s. \right\} \end{aligned} \tag{7.20}$$

consists of claims that can be superhedged without a cost by dynamic trading in the market.

Example 7.31. *Assume that $V = \delta_{\mathbb{R}_-}$ and let Assumption 7.28 hold. We then have $\bar{\varphi} = \delta_{\tilde{\mathcal{C}}}$ so, by Lemma 6.1, $\bar{\varphi}^* = \varphi^* = \sigma_{\tilde{\mathcal{C}}}$ and equation (7.17) becomes*

$$\sigma_{\tilde{\mathcal{C}}}(p, y) = E\left[\delta_{\mathbb{R}_+}(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(p_t + y\Delta s_{t+1})\right] + \delta_{\mathcal{X}_a^\perp}(p).$$

If the portfolio constraints D_t are conical, then σ_{D_t} is the indicator of the polar D_t° of D_t and $\varphi^* = \delta_{\tilde{\mathcal{C}}^\circ}$, where

$$\tilde{\mathcal{C}}^\circ = \{(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}_+ \mid p_t + y\Delta s_{t+1} \in D_t^\circ\}$$

is the polar of $\tilde{\mathcal{C}}$; see ?THM? **??**. *If $(p, y) \in \tilde{\mathcal{C}}^\circ$, then $E_t[y\Delta s_{t+1}] \in D_t^\circ$. If there are no portfolio constraints, we have $D_t = \mathbb{R}^J$ so that $D_t^\circ = \{0\}$ and the condition becomes $E_t[y\Delta s_{t+1}] = 0$ which, for nonzero $y \in \mathcal{Y}_+$ means that s is a martingale under the probability measure Q defined by $dQ/dP := y/Ey$; see ?THM?* **??**. *Similarly, if $D_t = \mathbb{R}_+^J$ (prohibition of short selling), the condition means that s is a supermartingale under Q.*

If $\tilde{\mathcal{C}}$ is closed, then by Theorem 7.31, the support function of $\mathcal{C}$ has the expression

$$\sigma_{\mathcal{C}}(y) = \delta_{\mathcal{C}}^*(y) = \bar{\varphi}(0, \cdot)^*(y) = E\left[\delta_{\mathbb{R}_+}(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])\right].$$

If, in addition, D_t are conical, we get

$$\mathcal{C}^\circ = \{y \in \mathcal{Y}_+ \mid E_t[y\Delta s_{t+1}] \in D_t^\circ \ t = 0, \ldots, T-1 \ a.s\}.$$

Again, if there are no portfolio constraints, the elements of $\mathcal{C}^\circ$ are nonnegative multiples of martingale densities. Sufficient conditions for the closedness of $\tilde{\mathcal{C}}$ will be given in Section **??**.

The following example is concerned with a variant of problem (3.4) where the objective is replaced by the "optimized certainty equivalent" discussed briefly in ?THM? **??**. Its proof is a direct application of Theorem 6.7 and Theorem 6.2.

Example 7.32 (Optimized certainty equivalent)**.** *Consider the problem*

$$\begin{aligned} &\text{minimize} \quad && \mathcal{V}(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}) \ \text{ over } \ x \in \mathcal{N} \\ &\text{subject to} \quad && x_t \in D_t \ t = 0, \ldots, T \quad a.s, \end{aligned} \tag{7.21}$$

where $\mathcal{V} : L^0 \to \overline{\mathbb{R}}$ is the optimized certainty equivalent *associated with the normal integrand V, i.e.*

$$\mathcal{V}(c) := \inf_{\alpha \in \mathbb{R}} \{\alpha + EV(c - \alpha)\}$$

see ?THM? **??**. *This fits the general duality framework with time* t *running from* -1 *to* T, $x_{-1} = \alpha$, $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$, $\bar{u} = c$ *and*

$$f(\hat{x}, u, \omega) = \alpha + \sum_{t=0}^{T} \delta_{D_t(\omega)}(x_t) + V(u - \alpha - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega)),$$

where $\hat{x} = (\alpha, x)$. *Indeed, problem* (1.2) *can then be written as*

$$\text{minimize} \quad Ef(\hat{x}, u) \ \text{ over } \ \hat{x} \in \hat{\mathcal{N}}, \tag{7.22}$$

where $\hat{\mathcal{N}} := L^0(\mathcal{F}_{-1}) \times \mathcal{N}$. *In addition to Assumption* 7.28, *we set* $\mathcal{X}_{-1} = L^\infty$ *and* $\mathcal{V}_{-1} = L^1$.

The dual problem can be written as

$$\begin{aligned} &\text{maximize } \ E[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(p_t + y\Delta s_{t+1})] \ \text{ over } (p, y) \in \mathcal{X}_a^\perp \in \mathcal{Y} \\ &\text{subject to } \ p_{-1} + y = 1 \quad a.s. \end{aligned} \tag{7.23}$$

If Assumption 7.28 *holds, the reduced dual becomes*

$$\begin{aligned} &\text{maximize} \qquad E\left[yc - V^*(y) - \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])\right] \ \text{ over } \ y \in \mathcal{Y} \\ &\text{subject to} \qquad Ey = 1. \end{aligned} \tag{7.24}$$

Note that the dual problems are the same as (7.18) *and* (7.19) *except that they have the additional constraint requiring* $Ey = 1$. *The optimum value of* (7.21) *equals that of* (7.23) *and* (7.24) *if and only if the function*

$$\bar{\varphi}(z, u) := \inf_{x \in L^0} \{Ef(\hat{x}, u) \mid \hat{x} - z \in \hat{\mathcal{N}}\}$$

is closed at $(0, c)$ *with respect to* $\sigma(\hat{\mathcal{X}} \times \mathcal{U}, \hat{\mathcal{V}} \times \mathcal{Y})$, *where* $\hat{\mathcal{X}} := \mathcal{X}_{-1} \times \mathcal{X}$ *and* $\hat{\mathcal{V}} := \mathcal{V}_{-1} \times \mathcal{V}$. *Sufficient conditions for the closedness of* $\bar{\varphi}$ *will be given in Section* **??**. *If* (7.21) *and* (7.24) *are feasible, then the following are equivalent:*

1. (α, x) *solves* (7.22), y *solves* (7.24) *and* $\inf(7.22) = \sup(7.24)$*;*
2. (α, x) *is feasible in* (7.22), y *is feasible in* (7.24) *and*

$$y \in \partial V(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \alpha),$$
$$E_t[y\Delta s_{t+1}] \in N_{D_t}(x_t) \quad t = 0, \ldots, T.$$

Given any $y \in \mathcal{Y}_+$ *with* $Ey = 1$, *?THM?* **??** *gives*

$$E_t[y\Delta s_{t+1}] = E_t[y]E_t^Q[\Delta s_{t+1}],$$

where the measure Q is defined by $dQ/dP = y$. Thus, by positive homogeneity of the support function and by ?THM? **??**, (7.24) *can be written as*

$$\underset{Q\in\mathcal{Q}^{\mathcal{Y}}}{\text{maximize}}\ E^Q[c] - EV^*(dQ/dP) - E^Q\left[\sum_{t=0}^{T-1}\sigma_{D_t}(E_t^Q[\Delta s_{t+1}])\right],$$

where $\mathcal{Q}^{\mathcal{Y}}$ is the set of probability measures with $dQ/dP \in \mathcal{Y}$. If there are no portfolio constraints, this can be written as

$$\text{maximize} \quad E^Q[c] - E[V^*(dQ/dP)] \quad \text{over } Q \in \mathcal{Q}_s^{\mathcal{Y}},$$

where $\mathcal{Q}_s^{\mathcal{Y}}$ is the set of martingale measures in $\mathcal{Q}^{\mathcal{Y}}$. When $V(u) = \delta_{\mathbb{R}_-}$, problem (7.21) *becomes the superhedging problem* (1.2) *and $V^* = \delta_{\mathbb{R}_+}$. When $V(u) = \exp(u) - 1$, the optimized certainty equivalent becomes the* entropic risk measure

$$\mathcal{V}(c) = \ln Ee^c.$$

In this case, $V^*(y) = y\log y - y + 1 + \delta_{\mathbb{R}_+}(y)$, so that

$$EV^*(dQ/dP) = E^Q \ln(dQ/dP),$$

the entropy *of Q relative to P. When $V(u) = \beta u^+$ for $\beta > 1$, the optimized certainty equivalent becomes the Conditional Value at Risk and $V^* = \delta_{[0,\beta]}$.*

Proof. The Lagrangian integrand becomes

$$\begin{aligned} l(\hat{x}, y, \omega) &= \inf_{u\in\mathbb{R}}\{f(\hat{x}, u, \omega) - uy\} \\ &= \alpha - y(\alpha + \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}(\omega)) + \sum_{t=0}^{T}\delta_{D_t(\omega)}(x_t) - V^*(y) \end{aligned}$$

and the conjugate of f,

$$\begin{aligned} f^*(v, y, \omega) &= \sup_{\hat{x}\in\mathbb{R}^n}\{\hat{x}\cdot v - l(\hat{x}, y, \omega)\} \\ &= V^*(y) + \delta_{\{0\}}(v_{-1} + y - 1) + \sum_{t=0}^{T-1}\sigma_{D_t(\omega)}(v_t + y\Delta s_{t+1}(\omega)). \end{aligned}$$

Since Ef is finite at the origin, Theorem 5.5 says that

$$\varphi^*(p, y) = \begin{cases} E[V^*(y) + \sum_{t=0}^{T-1}\sigma_{D_t}(p_t + y\Delta s_{t+1})] & \text{if } (p, y) \in \hat{\mathcal{X}}_a^\perp \times \mathcal{Y} \\ & \text{and } p_{-1} + y = 1, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\hat{\mathcal{X}}_a^\perp := \{p \in \hat{\mathcal{V}} \mid E[z\cdot p] = 0\ \forall z \in \hat{\mathcal{N}} \cap \hat{\mathcal{X}}\}$. Defining a mapping $\pi : \mathcal{Y} \to \hat{\mathcal{V}}$ by $\pi(y)_{-1} := Ey - y$ and $\pi(y)_t := E_t[y\Delta s_{t+1}] - y\Delta s_{t+1}$ for $t \geq 0$, Jensen's inequality in Theorem 4.13 gives, for any $(p,y) \in \hat{\mathcal{X}}_a^\perp \times \mathcal{Y}$,

$$\begin{aligned}\varphi^*(p,y) &= E[V^*(y) + \delta_{\{0\}}(p_{-1}+y-1) + \sum_{t=0}^{T-1} \sigma_{D_t}(p_{t+1} + y\Delta s_{t+1})] \\ &\geq E[V^*(y) + \delta_{\{0\}}(Ey-1) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}])] \\ &= \varphi^*(\pi(y), y).\end{aligned}$$

Thus, by Theorem 6.7, $\sup(7.24) = \sup_{(p,y)\in\hat{\mathcal{V}}\times\mathcal{Y}}\{\langle \bar{u}, y\rangle - \varphi^*(p,y)\}$, so the claim follows from Theorem 6.2. □

The next example studies the semi-static portfolio optimization problem from Example 1.6.

Example 7.33 (Semi-static hedging)**.** *Consider again problem* (1.3)

$$\begin{aligned}&\text{minimize } \; EV\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \bar{c}\cdot\bar{x} + S_0(\bar{x})\right) \text{ over } (x,\bar{x}) \in \mathcal{N}\times\mathbb{R}^{\bar{J}},\\ &\text{subject to} \qquad x_t \in D_t \;\; t = 0,\dots,T \;\; a.s.\end{aligned}$$

from Example 1.6 *and assume that the claim c as well as the components of $\bar{c}$ belong to $\mathcal{U}$. We start by rewriting the problem in the equivalent form*

$$\begin{aligned}&\text{minimize } EV\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \bar{c}\cdot\bar{x} + \alpha\right) \text{over} (x,\bar{x},\alpha) \in \mathcal{N}\times\mathbb{R}^{\bar{J}}\times\mathbb{R},\\ &\text{subject to} \qquad x_t \in D_t \quad t = 0,\dots,T \;\; a.s.\\ &\qquad\qquad\qquad\qquad S_0(\bar{x}) \leq \alpha.\end{aligned}$$

This fits the general duality framework with time t running from -1 to T, $\mathcal{F}_{-1} = \{\Omega,\emptyset\}$, $x_{-1} = (\bar{x},\alpha) \in \mathbb{R}^{\bar{J}}\times\mathbb{R}$, $\bar{u} = c$ and

$$\begin{aligned}f(\hat{x},u,\omega) = V&\left(u - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}(\omega) - \bar{c}(\omega)\cdot\bar{x} + \alpha, \omega\right)\\ &+ \sum_{t=0}^{T} \delta_{D_t(\omega)}(x_t,\omega) + \delta_{\operatorname{epi} S_0}(\bar{x},\alpha),\end{aligned}$$

where $\hat{x} = (x_{-1}, x)$. Indeed, since V is nondecreasing, $(\bar{x}, x)$ solves (1.3) *if and only if $\hat{x} = (\bar{x}, S_0(\bar{x}), x)$ solves*

$$\text{minimize} \quad Ef(\hat{x},u) \quad \hat{x}\in\hat{\mathcal{N}},$$

where $\hat{\mathcal{N}} := L^0(\Omega,\mathcal{F}_{-1},P;\mathbb{R}^{\bar{J}}\times\mathbb{R})\times\mathcal{N}$. In addition to Assumption 7.28*, we set $\mathcal{X}_{-1} = L^\infty$ and $\mathcal{V}_{-1} = L^1$.*

The optimum value of (1.3) *equals that of*

$$\underset{y\in\mathcal{Y}}{\text{maximize}} \quad E\left[cy - V^*(y) - \sum_{t=0}^{T-1}\sigma_{D_t}(E_t[y\Delta s_{t+1}])\right] - \sigma_{\text{epi}\, S_0}(E[y\bar{c}], -E[y]) \tag{7.25}$$

if and only if the function

$$\bar{\varphi}(z,u) := \inf_{x\in\ L^0}\{Ef(\hat{x},u) \mid \hat{x} - z \in \hat{\mathcal{N}}\}$$

is closed at $(0,c)$ *with respect to* $\sigma(\hat{\mathcal{X}}\times\mathcal{U}, \hat{\mathcal{V}}\times\mathcal{Y})$*, where* $\hat{\mathcal{X}} := \mathcal{X}_{-1}\times\mathcal{X}$ *and* $\hat{\mathcal{V}} := \mathcal{V}_{-1}\times\mathcal{V}$*. If* (1.3) *and* (7.25) *are feasible, then the following are equivalent:*

1. $(\bar{x},x)$ *solves* (1.3)*,* y *solves* (7.25) *and* $\inf(\textbf{??}) = \sup(7.25)$*;*
2. $(\bar{x},x)$ *is feasible in* (1.3)*,* y *is feasible in* (7.25) *and*

$$y \in \partial V\left(c - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1} - \bar{c}\cdot\bar{x} + S_0(\bar{x})\right),$$
$$E_t[y\Delta s_{t+1}] \in N_{D_t}(x_t) \quad t = 0,\dots,T,$$
$$E[y\bar{c}] \in \partial(E[y]S_0)(\bar{x})$$

almost surely.

Proof. The corresponding Lagrangian integrand becomes

$$\begin{aligned}
l(\hat{x},y,\omega) &= \inf_{u\in\mathbb{R}}\{f(\hat{x},u,\omega) - uy\} \\
&= \inf_{u\in\mathbb{R}}\left\{V\left(u - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}(\omega) - \bar{c}(\omega)\cdot\bar{x} + \alpha,\omega\right) - uy\right\} \\
&\quad + \sum_{t=0}^{T}\delta_{D_t(\omega)}(x_t,\omega) + \delta_{\text{epi}\, S_0}(\bar{x},\alpha) \\
&= y\left(\alpha - \bar{c}(\omega)\cdot\bar{x} - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}(\omega)\right) - V^*(y,\omega) \\
&\quad + \sum_{t=0}^{T}\delta_{D_t(\omega)}(x_t) + \delta_{\text{epi}\, S_0}(\bar{x},\alpha),
\end{aligned}$$

and the conjugate of f,

$$
\begin{aligned}
f^*(v,y,\omega) &= \sup_{\hat{x}\in\mathbb{R}^n}\{\hat{x}\cdot v - l(x,y,\omega)\}\\
&= V^*(y,\omega) + \sum_{t=0}^{T-1}\sup_{x_t\in\mathbb{R}^J}\{x_t\cdot(v_t + y\Delta s_{t+1}(\omega)) - \delta_{D_t(\omega)}(x_t)\}\\
&\qquad + \sup_{(\bar{x},\alpha)\in\mathbb{R}^{\bar{J}}\times\mathbb{R}}\{(\bar{x},\alpha)\cdot(v_{-1} + (y\bar{c}(\omega),-y) - \delta_{\operatorname{epi} S_0}(\bar{x},\alpha)\}\\
&= V^*(y,\omega) + \sum_{t=0}^{T-1}\sigma_{D_t(\omega)}(v_t + y\Delta s_{t+1}(\omega))\\
&\qquad + \sigma_{\operatorname{epi} S_0}(v_{-1} + (y\bar{c}(\omega),-y)).
\end{aligned}
$$

By Theorem 5.5,

$$
\varphi^*(p,y) = E\left[V^*(y) + \sum_{t=0}^{T-1}\sigma_{D_t}(p_t + y\Delta s_{t+1}) + \sigma_{\operatorname{epi} S_0}(p_{-1} + (y\bar{c},-y))\right]
$$

if $(p,y)\in\hat{\mathcal{X}}_a^\perp\times\mathcal{Y}$ and $\varphi^*(p,y)=\infty$ otherwise. Here

$$
\mathcal{X}_a^\perp := \{p\in\bar{\mathcal{V}} \mid E[z\cdot p]=0\ \forall z\in\hat{\mathcal{N}}\times\hat{\mathcal{X}}\}.
$$

Defining $\pi:\mathcal{Y}\to\hat{\mathcal{V}}$ by

$$
\begin{aligned}
\pi(y)_{-1} &:= (E[y\bar{c}] - y\bar{c}, -E[y] + y)\\
\pi(y)_t &:= E_t[y\Delta s_{t+1}] - y\Delta s_{t+1}\quad t=0,\dots,T-1,
\end{aligned}
$$

Jensen's inequality in Theorem 4.13 gives, for any $(p,y)\in\hat{\mathcal{X}}_a^\perp\times\mathcal{Y}_+$,

$$
\begin{aligned}
\varphi(p,y) &= E\left[V^*(y) + \sum_{t=0}^{T-1}\sigma_{D_t}(p_t + y\Delta s_{t+1}) + \sigma_{\operatorname{epi} S_0}(p_{-1} + (y\bar{c},-y))\right]\\
&\ge E\left[V^*(y) + \sum_{t=0}^{T-1}\sigma_{D_t}(E_t[y\Delta s_{t+1}]) + \sigma_{\operatorname{epi} S_0}(E[y\bar{c}],-E[y])\right]\\
&= \varphi(\pi(y),y).
\end{aligned}
\tag{7.26}
$$

Thus, the claims follow from Theorem 6.7 by choosing γ as the identity mapping. Indeed, the conditions

$$
\pi(y)\in\partial_x l(\hat{x},y),\quad \bar{u}\in\partial_y[-l](\hat{x},y)
$$

can be written as

$$
\begin{aligned}
(E[y\bar{c}],-E[y]) &\in N_{\operatorname{epi} S_0}(\bar{x},\alpha),\\
\pi(y)_t &\in -y\Delta s_{t+1} + N_{D_t}(x_t)\quad t=0,\dots,T,\\
c &\in -\left(\alpha - \bar{c}(\omega)\cdot\bar{x} - \sum_{t=0}^{T-1}x_t\cdot\Delta s_{t+1}(\omega)\right) + \partial V^*(y).
\end{aligned}
$$

By ?THM? **??**, the last condition can be written as

$$y \in \partial V\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \bar{c} \cdot \bar{x} + \alpha\right). \tag{7.27}$$

If $E[y] > 0$, then the first condition says that $\alpha = S_0(\bar{x})$ and

$$\frac{E[y\bar{c}]}{E[y]} \in \partial S_0(\bar{x})$$

which is equivalent to the last subdifferential condition in the statement. If $E[y] = 0$, then $y = 0$ and the first condition simply means that $S_0(\bar{x}) \le \alpha$. Since V is nondecreasing, (7.27) implies the first condition in the statement. If optimality conditions in the statement hold, the above conditions holds with $\alpha = S_0(\bar{x})$. □

The last term in the objective of (7.25) simplifies into a constraint in the case of a sublinear cost function S_0.

Remark 7.34 (Calibration of martingale measures)**.** *The support function of* epi S_0 *in the dual problem* (7.25) *in Example* 7.34 *can be expressed, by ?THM?* **??***, as*

$$\sigma_{\operatorname{epi} S_0}(v, -\alpha) = \begin{cases} \alpha S_0^*(v/\alpha) & \text{if } \alpha > 0, \\ (S_0^*)^\infty(v) & \text{if } \alpha = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

If S_0 is sublinear, then $S_0^*(\cdot, \omega)$ is the indicator of a closed convex set. Denoting this set by $C_0(\omega)$, we have

$$\sigma_{\operatorname{epi} S_0}(v, -\alpha) = \begin{cases} \delta_{C_0}(v/\alpha) & \text{if } \alpha > 0, \\ \delta_{C_0^\infty}(v) & \text{if } \alpha = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

For $\alpha \ge 0$, this equals $\delta_{\alpha_+ C_0}(v)$ where

$$\alpha_+ C_0 := \begin{cases} \alpha C_0 & \text{if } \alpha > 0, \\ C_0^\infty & \text{if } \alpha = 0. \end{cases}$$

The reduced dual (7.25) *can then be written with explicit constraints as*

$$\begin{aligned} &\text{maximize} && E\left[cy - V^*(y) - \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y \Delta s_{t+1}])\right] \quad \text{over } y \in \mathcal{Y} \\ &\text{subject to} && E[y\bar{c}] \in E[y]_+ C_0. \end{aligned}$$

If $E[y] > 0$, the constraint means that

$$E^Q \bar{c} \in C_0,$$

where Q is the probability measure defined by $dQ/dP = y/Ey$. The constraint thus requires that the measure Q be "calibrated" to the observed market prices of the claims $\bar{c}$. For example, if infinite quantities are available to buy and sell at prices $s^a \in \mathbb{R}^{\bar{J}}$ and $s^b \in \mathbb{R}^{\bar{J}}$, respectively (see Example 1.6), then $C_0 = [s^b, s^a]$ and the constraint can be written as the vector inequalities

$$s^b \le E^Q[\bar{c}] \le s^a$$

saying that the measure Q "prices" the claims between the bid-ask spread.

We end this section by some functional analytic reformulations of the studied problems.

Remark 7.35. *The optimum value of (7.21) can be expressed as*

$$\varphi_{OCE}(c) = \inf_{\alpha \in \mathbb{R}} \{\alpha + \varphi_0(c - \alpha)\},$$

where φ_0 is the optimum value function of problem Example 3.4 with respect to c. The conjugate of φ_{OCE} can thus be written as

$$\begin{aligned}
\varphi^*_{OCE}(y) &= \sup_{c,\alpha} \{\langle c, y\rangle + \alpha - \varphi_0(c - \alpha)\} \\
&= \sup_{c,\alpha} \{\langle c, y\rangle - \alpha - \varphi_0(c - \alpha)\} \\
&= \sup_{c,\alpha} \{\langle c', y\rangle - \alpha(1 - E[y]) - \varphi_0(c')\} \\
&= \begin{cases} \varphi^*_0(y) & \text{if } E[y] = 1, \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

If φ_{OCE} is closed and proper then, by Theorem 2.23,

$$\varphi_{OCE}(c) = \sup_y \{\langle c, y\rangle - \varphi^*_0(c) \mid E[y] = 1\}.$$

If the assumptions of Theorem 7.31 are satisfied, we thus get

$$\varphi_{OCE}(c) = \sup_{Q \in \mathcal{Q}^{\mathcal{Y}}} \{E^Q[c] - EV^*(dQ/dP) - E^Q \sum_{t=0}^{T-1} \sigma_{D_t}(E^Q_t[\Delta s_{t+1}])\}.$$

Note that if $V = \delta_{\mathbb{R}_-}$, then φ_{OCE} coincides with the function

$$\varphi_{SH}(c) := \inf_{\alpha \in \mathbb{R}} \{\alpha \mid c - \alpha \in \mathcal{C}\}$$

where $\mathcal{C}$ is the set of claims that can be superhedged without a cost; see (7.20). Under the above assumptions, we thus have the dual representation

$$\varphi_{SH}(c) = \sup_{Q \in \mathcal{Q}^{\mathcal{Y}}} \{E^Q[c] - E^Q \sum_{t=0}^{T-1} \sigma_{D_t}(E^Q_t[\Delta s_{t+1}])\},$$

where $\mathcal{Q}^{\mathcal{Y}}$ is the set of probability measures with $dQ/dP \in \mathcal{Y}$. If, in addition, there are no portfolio constraints, this becomes

$$\varphi_{SH}(c) = \sup_{Q \in \mathcal{Q}^{\mathcal{Y}}_s} E^Q[c],$$

where $\mathcal{Q}^{\mathcal{Y}}_s$ is the set of elements of $\mathcal{Q}^{\mathcal{Y}}$ under which s is a martingale.

The following reformulates the optimum value function of problem (1.3) in terms of that of (3.4).

Remark 7.36. *The optimum value of problem* (1.3) *can be expressed as*

$$\varphi_{SSH}(c) = \inf_{\bar{x} \in \operatorname{dom} S_0} \varphi_0(c + S_0(\bar{x}) - \bar{x} \cdot \bar{c}),$$

where φ_0 is the optimum value function of Example 3.4 *with respect to c. Thus, the conjugate of φ_{SSH} can be written as*

$$\begin{aligned}
\varphi^*_{SSH}(y) &= \sup_{c, \bar{x} \in \operatorname{dom} S_0} \{\langle c, y\rangle - \varphi_0(c + S_0(\bar{x}) - \bar{x} \cdot \bar{c})\} \\
&= \sup_{c', \bar{x} \in \operatorname{dom} S_0} \{\langle c' - S_0(\bar{x}) + \bar{x} \cdot \bar{c}, y\rangle - \varphi_0(c')\} \\
&= \varphi^*_0(y) + \sup_{\bar{x} \in \operatorname{dom} S_0} \{\langle \bar{x} \cdot \bar{c}, y\rangle - \langle S_0(\bar{x}), y\rangle\} \\
&= \varphi^*_0(y) + \sup_{\bar{x} \in \operatorname{dom} S_0} \{\bar{x} \cdot E[y\bar{c}] - E[y] S_0(\bar{x})\} \\
&= \varphi^*_0(y) + \sigma_{\operatorname{epi} S_0}(E[y\bar{c}], -E[y]) \\
&= \varphi^*_0(y) + \begin{cases} E[y] S^*_0(E[y\bar{c}]/E[y]) & \text{if } E[y] > 0, \\ (S^*_0)^\infty(E[y\bar{c}]) & \text{if } E[y] = 0, \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Much as in Remark 7.36*, one can then use Theorem* 2.23 *and the dual representation of φ_0 in Theorem* 7.31 *to derive a dual representation for φ_{SSH}.*

7.6 Subdifferentials and conditional expectations

This section applies the results of Section 5 to the general theory of conditional expectations of normal integrands. More precisely, Theorem 7.38 below gives sufficient conditions for the commutation of conditional expectation and subdifferentiation of a normal integrand. The result is interesting in its own right but it will not be needed for subsequent developments in this book. Sufficient conditions for the assumptions of Theorem 7.38 will be given in Section **??**.

Given a convex normal integrand h and an $x \in L^0(\mathbb{R}^n)$, Theorem 3.27 says that the set valued mapping $\partial h(x)(\omega) := \partial h(x(\omega), \omega)$ is closed convex-valued and measurable. We denote the $\mathcal{G}$-measurable selections of a set-valued mapping $S : \Omega \rightrightarrows \mathbb{R}^n$ by $\mathcal{X}(\mathcal{G}; S)$ or, if $S = \mathbb{R}^n$ almost surely, simply $\mathcal{X}(\mathcal{G})$. Recall from

?THM? **??**, that if S is closed convex-valued and measurable, then it has a $\mathcal{G}$-conditional expectation for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. The $\mathcal{G}$-conditional expectation is the unique closed convex-valued $\mathcal{G}$-measurable mapping $E^{\mathcal{G}} S$ such that

$$L^1(\mathcal{G}; E^{\mathcal{G}} S) = \mathrm{cl}\{E^{\mathcal{G}} v \mid v \in L^1(\mathcal{F}; S)\},$$

where the closure is taken in the L^1-topology.

Theorem 7.37. *Let h be a convex normal integrand and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra such that Eh^* is proper on $\mathcal{V}$ and, for all $v \in \mathcal{V}$, the function $\phi_v : \mathcal{X} \to \overline{\mathbb{R}}$ given by*

$$\phi_v(z) := \inf_{x \in \mathcal{X}} \{E[h(x) - x \cdot v] \mid x - z \in \mathcal{X}(\mathcal{G})\}$$

is either subdifferentiable or equal to $-\infty$ at the origin. We have

$$\mathcal{V}(\mathcal{G}; E^{\mathcal{G}}[\partial h(x)]) = \mathcal{V}(\mathcal{G}; \partial(E^{\mathcal{G}} h)(x))$$

for all $x \in \mathrm{dom}\, Eh \cap \mathcal{X}(\mathcal{G})$. If there exists an $\bar{x} \in \mathrm{dom}\, Eh \cap \mathcal{X}(\mathcal{G})$ such that $\mathcal{V}(\partial h(\bar{x})) \neq \emptyset$, then

$$\partial(E^{\mathcal{G}} h)(x) = E^{\mathcal{G}}[\partial h(x)] \quad a.s.$$

for every $x \in L^0(\mathcal{G})$ such that $h(x) \in L^1$ and $\mathcal{V}(\partial h(x)) \neq \emptyset$.

Proof. Let $x \in \mathrm{dom}\, Eh \cap \mathcal{X}(\mathcal{G})$. By Fenchel's inequality, $h(x) + h^*(v) \geq x \cdot v$, so the properness of Eh^* on $\mathcal{V}$ implies that h is L-bounded and $h(x)^- \in L^1$. Thus, by ?THM? **??**,

$$E^{\mathcal{G}} v \in \partial(E^{\mathcal{G}} h)(x) \ a.s.$$

for every $v \in \mathcal{V}(\partial h(x))$. Since L^1-convergence implies almost sure convergence and since $\partial(E^{\mathcal{G}} h)(x)$ is closed-valued, we thus get

$$L^1(\mathcal{G}; E^{\mathcal{G}}[\partial h(x)]) \subseteq L^1(\mathcal{G}; \partial(E^{\mathcal{G}} h)(x))$$

and, in particular,

$$\mathcal{V}(\mathcal{G}; E^{\mathcal{G}}[\partial h(x)]) \subseteq \mathcal{V}(\mathcal{G}; \partial(E^{\mathcal{G}} h)(x)).$$

To prove the converse, let $v \in \mathcal{V}(\mathcal{G}; \partial(E^{\mathcal{G}} h)(x))$. Since h is L-bounded and $h(x)^- \in L^1$, Theorem 3.29 and ?THM? **??** give

$$\begin{aligned}
\phi_v(0) &= \inf_{x' \in \mathcal{X}(\mathcal{G})} E[h(x') - x' \cdot v] \\
&= \inf_{x' \in \mathcal{X}(\mathcal{G})} E[(E^{\mathcal{G}} h)(x') - x' \cdot v] \\
&= E[\inf_{x' \in \mathbb{R}^n} \{(E^{\mathcal{G}} h)(x') - x' \cdot v\}] \\
&= E[(E^{\mathcal{G}} h)(x) - x \cdot v] \\
&= E[h(x) - x \cdot v],
\end{aligned}$$

which is finite since $x \in \mathrm{dom}\, Eh$ and $h(x)^- \in L^1$.

We apply Theorem 5.3 to the one-period instance of $(P_{\mathcal{X}})$ where $T = 0$, $\mathcal{F}_0 = \mathcal{G}$, $\bar{u} = 0$ and

$$f(x, u, \omega) := h(x, \omega) - x \cdot v.$$

The assumption on ϕ_v implies $\phi_v(0) < \infty$ so $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ and, by Theorem 5.5,

$$F^*(v, p, y) = Ef^*(v + p, y) + \delta_{\mathcal{X}_a^\perp}(p)$$

and, in particular,

$$\varphi^*(p, y) = Ef^*(p, y) + \delta_{\mathcal{X}_a^\perp}(p),$$

where $f^*(p, y, \omega) = h^*(v + p, \omega) + \delta_{\{0\}}(y)$ and $\mathcal{X}_a^\perp = \{p \in \mathcal{V} \mid E^{\mathcal{G}} p = 0\}$. Thus, by Theorem 5.3,

$$\inf_{x \in \mathcal{X}(\mathcal{G};\mathbb{R}^n)} E[h(x) - x \cdot v] = \sup_{p \in \mathcal{V}} \{-Eh^*(v + p) \mid E^{\mathcal{G}} p = 0\},$$

where the supremum is attained by some $\bar{p} \in \mathcal{V}$. Since Eh is proper on $\mathcal{X}(\mathcal{G})$, ?THM? **??** and the interchange rule in Theorem 3.29 give

$$\begin{aligned} \inf_{x \in \mathcal{X}(\mathcal{G};\mathbb{R}^n)} E[h(x) - x \cdot v] &= \inf_{x \in \mathcal{X}(\mathcal{G};\mathbb{R}^n)} \{E(E^{\mathcal{G}} h)(x) - \langle x, v \rangle\} \\ &= -E(E^{\mathcal{G}} h)^*(v) \end{aligned}$$

so $Eh^*(v + \bar{p}) = E(E^{\mathcal{G}} h)^*(v)$. By ?THM? **??** and ?THM? **??**, $E^{\mathcal{G}} h^*(v + \bar{p}) \geq (E^{\mathcal{G}} h)^*(v)$ almost surely, so we must have,

$$E^{\mathcal{G}} h^*(v + \bar{p}) = (E^{\mathcal{G}} h)^*(v) \quad a.s. \tag{7.28}$$

Since $v \in \partial(E^{\mathcal{G}} h)(x)$ almost surely,

$$(E^{\mathcal{G}} h)(x) + (E^{\mathcal{G}} h)^*(v) = x \cdot v \quad a.s.$$

so, by (7.28),

$$(E^{\mathcal{G}} h)(x) + E^{\mathcal{G}} h^*(v + \bar{p}) = x \cdot v \quad a.s.$$

Since $E[x \cdot \bar{p}] = 0$ and $h(x)$ is integrable, ?THM? **??** gives

$$E[h(x) + h^*(v + \bar{p}) - x \cdot (v + \bar{p})] = 0.$$

This implies that the Fenchel's inequality $h(x) + h^*(v + \bar{p}) - x \cdot (v + \bar{p}) \geq 0$ must hold as an equality almost surely. The equality means that $v + \bar{p} \in \partial h(x)$. Taking conditional expectations, we get $v \in E^{\mathcal{G}}[\partial h(x)]$ almost surely so

$$v \in \mathcal{V}(\mathcal{G}; E^{\mathcal{G}}[\partial h(x)])$$

which completes the proof of the first claim.

Let $x \in L^0(\mathcal{G})$ such that $h(x) \in L^1$ and $\mathcal{V}(\partial h(x)) \neq \emptyset$. Define, for every $\nu \in \mathbb{N}$, $x^\nu := 1_{A^\nu} x + 1_{\Omega \setminus A^\nu} \bar{x}$, where $A^\nu = \{|x| \leq \nu\}$. We have $x^\nu \in \operatorname{dom} Eh \cap \mathcal{X}(\mathcal{G})$ and $\mathcal{V}(\partial h(x^\nu)) \neq \emptyset$. By ?THM? **??**, the first claim implies

$$\partial(E^{\mathcal{G}} h)(x^\nu) = E^{\mathcal{G}}[\partial h(x^\nu)] \quad a.s.$$

By ?THM? **??** and ?THM? **??**.1,

$$E^{\mathcal{G}}[\partial h(x^\nu)] = \begin{cases} E^{\mathcal{G}}[\partial h(x)] & \text{on } A^\nu, \\ E^{\mathcal{G}}[\partial h(\bar{x})] & \text{on } \Omega \setminus A^\nu. \end{cases}$$

In particular, $\partial(E^{\mathcal{G}} h)(x) = E^{\mathcal{G}}[\partial h(x)]$ on A^ν. It now suffices to note that $\bigcup_\nu A^\nu = \Omega$. □