Convex Stochastic optimization WiSe 2020/2021

Ari-Pekka Perkkiö

February 1, 2021

Contents

1 Practicalities and background

Upon passing the exam, attending and solving the exercises give a bonus to the final grade.

We assume that the following concepts are familiar:

1. Vector spaces, topology.
2. Probability space, random variables, expectation, convergence theorems.
3. Conditional expectations, martingales.
4. The fundamentals of discrete time financial mathematics.

2 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

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

where, for given integers n_t and m,

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

is the space of *adapted processes*, h is an extended real-valued $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable function, where $n := n_0 + \ldots + n_T$. 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 $\omega \in \Omega$. Then Eh is a convex function of $\mathcal{N}$ and (SP$_u$) is a *convex stochastic optimization problem* on the space of adapted processes.

The aim of this course is to analyze (SP$_u$) using *dynamic programming* and *conjugate duality*. These lead to characterizations of optimal solutions of (SP$_u$) 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$_u$), a classical example being that the superhedging price of an option in a liquid market can be computed via "martingale measures".

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

2.1 Examples

Example 2.1 (Mathematical programming). *Consider the problem*

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

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

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

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 2.2 (Optimal stopping). *Consider the problem*

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

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

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

Example 2.3 (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*

$$\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}_D, \tag{2.1}$$

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

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

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.*

Example 2.4 (Stochastic control)**.** *The problem*

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

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 2.5 (Problems of Bolza)**.** *Consider the problem*

$$\underset{x\in\mathcal{N}}{\text{minimize}} \quad E\sum_{t=0}^{T} K_t(x_t, \Delta x_t), \tag{2.2}$$

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

Consider the problem

$$\underset{x \in \mathcal{N}}{\text{minimize}} \quad \mathcal{V}(x) \tag{2.3}$$

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

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

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

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. □

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. $\omega \mapsto (S^1(\omega), \dots, S^j(\omega))$ *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*

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

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

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

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

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

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

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 $\beta \in \mathbb{R}$, $\operatorname{lev}_{\le\beta} h$ is closed-valued and measurable by Theorem 3.7, so $\{(x,\omega) \mid h(x,\omega) \le \beta\}$ 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).

□

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 8.12 in the appendix gives the following.

Corollary 3.13. *Let h be a normal integrand such that $h(x) \geq -\rho|x| - m$ 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 $h^\nu(x) \geq -\rho|x| - m$ 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. By Theorem 3.14, it suffices to show that the necessity is equivalent to the existence of a Castaing representation. If a Castaing representation, exists, it satisfies 1 and 2. When (x^ν) satisfies 1 and 2,...?? □

Theorem 3.19. *For normal integrands h and $\tilde{h}$, we have $h \le \tilde{h}$ if and only if, for every $w \in \mathcal{L}^\infty(\mathbb{R}^n)$, $h(w) \le \tilde{h}(w)$. 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 $h \ge -m$ 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 $h(w) \le \tilde{h}(w)$ if and only if $h^\nu(w) \le \tilde{h}^\nu(w)$ 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

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

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, $w := \hat{w}\mathbb{1}_{|\hat{w}| \le \nu}$ and $A := \{|\hat{w}| \le \nu\} \cap \operatorname{dom} S$ is nonempty which is a contradiction with the assumption that $h(w) \le \tilde{h}(w)$. □

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. TBA □

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 8.18, 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 8.18, 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

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

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* $(-K) \cap \{u \in \mathbb{R}^m \mid g^\infty(u, \omega) \leq 0\}$ *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 $K = \{0\}$. □

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. □

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 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^ν) and $x \in L^0$. If $x^\nu \to x$ in probability, there is a subsequence such that $P(|x^{\nu'} - x|) \le 2^{-\nu}$ for all $\nu \ge \nu'$. Thus $E[\sum_{\nu'=1}^{\infty} 1_{|x^{\nu'} - x|}] \le 1$ by monotone convergence, so $\sum_{\nu'=1}^{\infty} 1_{|x^{\nu'} - x|} < \infty$ almost surely, implying $x^{\nu'} \to x$ 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 $P(|x^{\nu'} - x|) > \epsilon$. Taking a subsequence, dominated convergence gives a contradiction.

If $x^\nu \to x$ in probability, then, taking a almost surely converging subsequence, dominated convergence implies that $d(x^\nu, x) \to 0$. If x^ν does not converge to x in probability, there is an $\epsilon > 0$, $\delta > 0$ and a subsequence such that $P(|x^{\nu'} - x| \ge \epsilon) \ge \delta$. Then $d(x^{\nu'}, x) \ge \delta\rho(\epsilon)$, so subsequences of $(x^{\nu'})$ cannot converge to x.

Clearly, L^0 is a complete metric vector space while (sequential) continuity of addition and multiplication follow from the dominated convergence theorem. □

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 bounded from below.*
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 8.5.1, $\bar{x} \in C^\infty$, 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. □

The following gives sufficient conditions for existence of solutions. The conditions will be generalized first in Theorem 3.42 below and later in Chapter **??** after the development of the dynamic programming principle.

Theorem 3.38. *Assume that h is a convex normal integrand bounded from below such that*

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

almost surely. Then (SP$_u$) *has an optimal solution.*

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

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

Komlos theorem (Lemma 8.33) 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

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

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

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

Theorem 8.17 gives

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

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

By Lemma 8.33, 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, $Ek(\hat{x}^\nu) \to \inf(\mathrm{SP}_u)$. The function Ek is lsc on L^0 by Theorem 3.31, so

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

which completes the proof. □

Clearly, the second condition in Theorem 3.38 holds if, for P-almost every $\omega \in \Omega$,

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

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

The lower boundedness in Theorem 3.38 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 3.39. *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 3.40. *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 3.39 with $v \in \mathcal{N}^\perp$ defined by $v_t = \Delta s_{t+1}$.

Definition 3.41. *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 3.42. *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_u) 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 3.38 proves the claim. □

Corollary 3.43. *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_u) 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,

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

Thus the claim follows from Theorem 3.38. □

Lemma 3.44. *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

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

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

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

Let $x \in \mathcal{N}$. If either $Eh(x) < \infty$ or $E[h(x) - x \cdot p] < \infty$, the above inequalities and Lemma 3.39 give $E[x \cdot p] = 0$, so

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

The above inequalities also give

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

If either $h^\infty(x,0) \leq 0$ or $h^\infty(x,0) - x \cdot p \leq 0$ almost surely, then $x \cdot p \leq 0$ almost surely. Lemma 3.39 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 3.44 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 3.45 (Mathematical programming)**.** *Consider the problem*

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

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

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

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

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

then the problem has a solution as soon as

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

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

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

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

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

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 3.38 □

Example 3.46 (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 3.45.*

In the linear case, Example 3.45 can be written as follows.

Example 3.47 (Linear programming)**.** *Consider the problem*

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

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

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

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

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

3.7.2 Stochastic control and problems of Bolza

Example 3.48 (Stochastic control)**.** *Consider the problem*

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

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 3.38. □

Example 3.49 (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

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

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 3.38. □

3.7.3 Financial mathematics

Later, we formulate optimal investment problem as stochastic control.

Example 3.50 (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

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

and the recession condition in Theorem 3.38 becomes the last condition in the statement. □

Example 3.51 (Semistatic hedging)**.** *Consider the problem*

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

Proof. The problem fits to the general framework as above by extending x_0 to $(z_0, \bar{z}_0)$. □

Example 3.52 (Currency markets)**.** *Consider Example 3.49 with*

$$K_t(x_{t-1}, \Delta x_t) = \delta_{D_{t-1}}(x_{t-1}, \omega) + \delta_{C_t}(\Delta x_t, \omega)$$

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 Dynamic programming

The purpose of this section is to prove a general *dynamic programming* recursion which generalizes the classical Bellman equation for convex stochastic optimization. We will use the notion of a conditional expectation of a normal integrand. In certain financial applications, the new condition turns out to be equivalent to the classical no-arbitrage condition.

An extended real-valued random variable X is said to *quasi-integrable* if either X^+ or X^- is integrable. Given a quasi-integrable X and a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists an extended real-valued $\mathcal{G}$-measurable random variable $E^{\mathcal{G}}X$, unique almost surely, such that

$$E\left[\alpha(E^{\mathcal{G}}X)\right] = E\left[\alpha X\right] \quad \forall \alpha \in L^\infty_+(\Omega, \mathcal{G}, P).$$

The random variable $E^{\mathcal{G}}X$ is known as the $\mathcal{G}$-conditional expectation of X.

Definition 4.1. *Given a normal integrand h, a $\mathcal{G}$-normal integrand $E^{\mathcal{G}}h$ is a* $\mathcal{G}$-conditional expectation of h *if*

$$(E^{\mathcal{G}}h)(x) = E^{\mathcal{G}}[h(x)] \quad a.s.$$

for all $x \in L^0(\mathcal{G})$ such that $h(x)$ is quasi-integrable.

We will use notations $x^t = (x_0, \dots, x_t)$, $n^t = n_0 + \cdots + n_t$ and $E_t = E^{\mathcal{F}_t}$. For $t = -1$, x^t is understood as the vector with dimension zero. We say that an adapted sequence $(h_t)_{t=0}^T$ of normal integrands $h_t : \mathbb{R}^{n^t} \times \Omega \to \overline{\mathbb{R}}$ solves the generalized *Bellman equations for h* if

$$\begin{aligned} \tilde{h}_T &:= h, \\ h_t &:= E_t \tilde{h}_t, \\ \tilde{h}_{t-1}(x^{t-1}, \omega) &:= \inf_{x_t \in \mathbb{R}^{n_t}} h_t(x^{t-1}, x_t, \omega) \end{aligned} \tag{BE}$$

for $t = T, \dots, 0$. For the second equation to be well-defined, $\tilde{h}_t$ has to be a normal integrand.

In the above formulation, one does not separate the decision variables x_t into "state" and "control" like in the classical dynamic programming models. Formulations closer to the classical dynamic programming equations will be obtained as special cases of (BE) in Section 4.3.4 below.

Just as in the classical formulations, a solution (h_t) to (BE) provides optimality conditions for (SP_u). We will denote the projection of $\mathcal{N}$ to its first t components by $\mathcal{N}^t$. That is,

$$\mathcal{N}^t = \left\{ (x_{t'})_{t'=0}^t \;\middle|\; x_{t'} \in L^0(\Omega, \mathcal{F}_{t'}, P; \mathbb{R}^{n_{t'}}) \right\}.$$

Theorem 4.2. *Assume that h is bounded from below,* (SP_u) *is feasible and that the Bellman equations* (BE) *admit a solution $(h_t)_{t=0}^T$. Then*

$$\inf (\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} E h_t(x^t)$$

for all $t = 0, \ldots, T$ and an $\bar{x} \in \mathcal{N}$ solves (SP_u) *if and only if*

$$\bar{x}_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^{n_t}} h_t(\bar{x}^{t-1}, x_t) \quad a.s. \quad t = 0, \ldots, T. \tag{4.1}$$

Proof. Let $x \in \mathcal{N}$. Theorem 3.29 gives, for any t,

$$E\tilde{h}_{t-1}(x^{t-1}) = \inf_{x_t \in L^0(\mathcal{F}_t)} E h_t(x^{t-1}, x_t).$$

Since $\tilde{h}_{t-1}$ is bounded from below, $E\tilde{h}_{t-1} = E h_{t-1}$ on $\mathcal{N}^{t-1}$ and thus

$$\inf_{x^{t-1} \in \mathcal{N}^{t-1}} E h_{t-1}(x^{t-1}) = \inf_{x^t \in \mathcal{N}^t} E h_t(x^t).$$

By induction, $\inf (\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} E h_t(x^t)$.

To prove the second claim, note first that

$$\bar{x}^t \in \operatorname*{argmin}_{x^t \in \mathcal{N}^t} E h_t(x^t)$$

if and only if

$$\bar{x}^{t-1} \in \operatorname*{argmin}_{x^{t-1} \in \mathcal{N}^{t-1}} E h_{t-1}(x^{t-1}) \quad \text{and} \quad \bar{x}_t \in \operatorname*{argmin}_{x_t \in L^0(\mathcal{F}_t)} E h_t(\bar{x}^{t-1}, x_t),$$

where, by the second part of Theorem 3.29, the second inclusion means that

$$\bar{x}_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^{n_t}} h_t(\bar{x}^{t-1}, x_t) \quad \text{a.s.}$$

An $\bar{x} \in \mathcal{N}$ solves (SP_u) if and only if $\bar{x}$ minimizes $E h_T$. A backward recursion shows that optimal solutions satisfy (4.1). The converse follows from a forward recursion. □

Exercise 4.0.1. *Let $T = 1$, $n_0 = n_1 = 1$, ξ_1, ξ_2 be i.i.d. with $P(\xi_t = 1) = P(\xi_t = -1/2) = 1/2$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \sigma(\xi_1)$, $\mathcal{F} = \sigma(\xi_1, \xi_2)$, and*

$$h(x, \omega) = \exp(\sum_{t=0}^{1} x_t \xi_{t+1}(\omega)).$$

Find a solution (h_t) to (BE) *and an optimal solution of*

$$\inf_{x \in \mathcal{N}} Eh(x).$$

Hint: Let η^i be real-valued random variable with $\eta^1 \in L^0(\mathcal{G})$. If η^i are nonnegative, then

$$E^{\mathcal{G}}[\eta^1 \eta^2] = \eta^1 E^{\mathcal{G}}[\eta^2].$$

If η^2 is independent of $\mathcal{G}$ and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ is measurable, then

$$E^{\mathcal{G}}[g(\eta^1, \eta^2)] = E[g(x, \eta^2)]_{|x=\eta^1}. \tag{4.2}$$

These formulas will be justified later in more generality.

Proof. Note first that h is a Caratheodory integrand, and, in particular, a normal integrand. Since h is nonnegative, $h(x)$ is quasi-integrable for any $x \in L^0(\mathcal{F})$. Given $x \in L^0(\mathcal{F}_1)$,

$$\begin{aligned} E_1[h(x)] &= E_1[\exp(\sum_{t=0}^{1} x_t \xi_{t+1})] \\ &= \exp(x_0 \xi_1) E_1[\exp(x_1 \xi_2)] \\ &= \exp(x_0 \xi_1)[\frac{1}{2} \exp(x_1) + \exp(-x_1/2)], \end{aligned}$$

so

$$h_1(x^1, \omega) = \frac{1}{2} \exp(x_0 \xi_1(\omega))[\exp(x_1) + \exp(-x_1/2)]$$

and

$$\begin{aligned} \tilde{h}_0(x_0, \omega) &= \inf_{x_1} h_1(x_0, x_1, \omega) \\ \{\text{optimal } x_1 = \ln(2^{-2/3}\} &= \frac{1}{2}[2^{-2/3} + 2^{1/3}] \exp(x_0 \xi_1(\omega)) \\ &= [2^{-5/3} + 2^{-2/3}] \exp(x_0 \xi_1(\omega)) \end{aligned}$$

Similarly ($\mathcal{F}_0$ is trivial, so h_0 is deterministic)

$$h_0(x_0) = [2^{-5/3} + 2^{-2/3}] \frac{1}{2}[\exp(x_0) + \exp(-x_0/2)].$$

Now the optimal solution is

$$\bar{x}_0(\omega) \in \operatorname*{argmin}_{x_0} h_0(x_0) = \ln(2^{-2/3}), \quad \bar{x}_1(\omega) \in \operatorname*{argmin}_{x_1} h_1(\bar{x}_0, x_1, \omega) = \ln(2^{-2/3})$$

and the optimal value is

$$Eh_0(\bar{x}_0) = h_0(\bar{x}_0) = [2^{-5/3} + 2^{-2/3}]^2.$$

□

A solution (h_t) of (BE) is sometimes referred to as the *cost-to-go* function. This term is well motivated by the expression in Remark 4.3 below. Given a collection $C \subset L^0$ of random variables, a random variable $\xi \in L^0$ is said to be the *essential infimum* of C if $\xi \le \xi'$ almost surely for every $\xi' \in C$ and if $\underline{\xi} \in L^0$ is another random variable with this property, then $\underline{\xi} \le \xi$ almost surely. Every set C of random variables has a unique essential infimum; see Appendix **??**. We denote it by

$$\text{ess inf}_{\xi' \in C}\, \xi'.$$

Remark 4.3. *In the setting of Theorem 4.2,*

$$h_t(x^t) = \text{ess inf}_{\tilde{x} \in \mathcal{N}} \left\{ E_t h(\tilde{x}) \;\middle|\; \tilde{x}^t = x^t \right\} \quad \forall x^t \in L^0(\mathcal{F}_t).$$

Proof. Exercise. Hint: Show first that $C = \{\xi'_{\alpha,\alpha'} \mid \alpha \in \mathcal{J}, \alpha' \in \mathcal{J}'\}$, then

$$\text{ess inf}_{\xi' \in C}\, \xi' = \text{ess inf}_{\alpha \in \mathcal{J}} (\text{ess inf}_{\alpha' \in \mathcal{J}'}\, \xi'_{\alpha,\alpha'});$$

use also Lemma 4.4 below.

Given $x^t \in L^0(\mathcal{F}_t)$, the definition of (h_t) and Lemma 4.4 below give

$$\begin{aligned} h_t(x^t) &= E_t[\inf_{x_{t+1}} h_{t+1}(x^t, x_{t+1})] \\ &= \text{ess inf}_{x_{t+1} \in L^0(\mathcal{F}_{t+1})}\, E_t[h_{t+1}(x^t, x_{t+1})] \\ &= \text{ess inf}_{x_{t+1} \in L^0(\mathcal{F}_{t+1})}\, \text{ess inf}_{x_{t+2} \in L^0(\mathcal{F}_{t+2})}\, E_t h_{t+2}(x^t, x_{t+1}, x_{t+2}). \end{aligned}$$

It suffices to use the hint: if $C = \{\xi'_{\alpha,\alpha'} \mid \alpha \in \mathcal{J}, \alpha' \in \mathcal{J}'\}$, then

$$\text{ess inf}_{\xi' \in C}\, \xi' = \text{ess inf}_{\alpha \in \mathcal{J}} (\text{ess inf}_{\alpha' \in \mathcal{J}'}\, \xi'_{\alpha,\alpha'}),$$

so the statement follows from recursion. □

Lemma 4.4. *Assume that h is a normal integrand bounded from below. We have*

$$E^{\mathcal{G}}[\inf_{u \in \mathbb{R}^m} h(u)] = \text{ess inf}_{u \in L^0(\mathcal{F})}\, E^{\mathcal{G}}[h(u)]$$

Proof. Exercise. Hint: Mimick the proof of the interchange rule.

Clearly,

$$E^{\mathcal{G}}[\inf_{u \in \mathbb{R}^m} h(u)] \le \text{ess inf}_{u \in L^0(\mathcal{F})}\, E^{\mathcal{G}}[h(u)].$$

Let $\epsilon > 0$ and

$$S(\omega) := \{u' \in \mathbb{R}^m \mid h(u', \omega) \le \inf_{u \in \mathbb{R}^m} h(u, \omega) + \epsilon\}$$

Since h is a normal integrand, S is measurable closed-valued, so, by the measurable selection theorem (Corollary 3.15), there exists $\bar{u} \in L^0(\mathcal{F})$ such that $h(\bar{u}) \leq \inf_{u\in\mathbb{R}^m} h(u,\omega) + \epsilon$ almost surely. Thus

$$E^{\mathcal{G}} h(\bar{u}) \leq E^{\mathcal{G}}[\inf_{u\in\mathbb{R}^m} h(u)] + \epsilon$$

and

$$\text{ess}\inf_{u\in L^0(\mathcal{F})} E^{\mathcal{G}}[h(u)] \leq E^{\mathcal{G}}[\inf_{u\in\mathbb{R}^m} h(u)] + \epsilon$$

almost surely. Since $\epsilon > 0$ was arbitrary, this completes the proof. □

In Sections 4.1 and 4.2, we study existence of solutions to (BE) and (SP$_u$) and extend Theorem 4.2 to settings where the objective h is not bounded from below.

4.1 Conditional expectation of a normal integrand

Note that if $\tilde{h}$ is a $\mathcal{G}$-conditional expectation of h, then any $\bar{h}$ that is indistinguishable from $\tilde{h}$ is also a $\mathcal{G}$-conditional expectation of h. Uniqueness and equality of normal integrands means indistinguishable.

For an integrable $\mathbb{R}^d$-valued random variable X, $E^{\mathcal{G}}X$ is defined componentwise as the conditional exectation of X.

4.1.1 Existence

We say that a normal integrand h is *L-bounded* if there exist $\rho, m \in L^1$ such that

$$h(x) \geq -\rho|x| - m \quad \forall x \in \mathbb{R}^n. \tag{L}$$

Lemma 4.5. *For a convex normal integrand h, the following are equivalent:*

1. *h is L-bounded,*
2. *there exists $v \in L^1(\mathbb{R}^n)$ and $m \in L^1$ such that*
$$h(x,\omega) \geq x \cdot v(\omega) - m(\omega),$$
3. $\operatorname{dom} Eh^* \cap L^1 \neq \emptyset$.

Proof. Let $v \in L^1$ such that $Eh^*(v) < \infty$ is finite. By Fenchel's inequality,

$$h(x,\omega) \geq x \cdot v - h^*(v,\omega) \geq -|v||x| - h^*(v,\omega),$$

so we may choose $\rho = |v|$ and $m = h^*(v)^+$. On the other hand, $h \geq -\rho|\cdot| - m$ can be written as $(h + \rho|\cdot|)^*(0) \leq m$. By Corollary **??** in the appendix, this means that

$$\inf_{v\in\mathbb{R}^n}\{h^*(v) + \delta_{\mathbb{B}}(v/\rho)\} \leq m,$$

where the infimum is attained. By Corollary 3.15, there is a $v \in L^0$ with $|v| \leq \rho$ and $h^*(v) \leq m$. $\square$

For an L-bounded normal integrand, $E^{\mathcal{G}}[h(x)]$ is well-defined for every $x \in L^\infty(\mathcal{G})$. The following lemma shows that, in the definition of conditional expectation, for L-bounded normal integrands, it suffices to test with $x \in L^\infty(\mathcal{G})$.

Lemma 4.6. *Given an L-bounded normal integrand h, a $\mathcal{G}$-normal integrand $\tilde{h}$ is the unique $\mathcal{G}$-conditional expectation of h if*

$$\tilde{h}(x) = E^{\mathcal{G}}[h(x)]$$

almost surely for all $x \in L^\infty(\mathcal{G})$.

Proof. If there is $x \in L^0(\mathcal{G})$ such that $h(x)$ is quasi-integrable and $\tilde{h}(x) \neq E^{\mathcal{G}}[h(x)]$, then, for ν large enough,

$$\mathbb{1}_{\{|x|\leq\nu\}}\tilde{h}(\mathbb{1}_{\{|x|\leq\nu\}}x) \neq \mathbb{1}_{\{|x|\leq\nu\}}E^{\mathcal{G}}[h(\mathbb{1}_{\{|x|\leq\nu\}}x)] = E^{\mathcal{G}}[\mathbb{1}_{\{|x|\leq\nu\}}h(\mathbb{1}_{\{|x|\leq\nu\}}x)]$$

which is a contradiction. Uniqueness follows from Theorem 3.28. $\square$

Exercise 4.1.1. *Let $h(x,\omega) = x \cdot v(\omega)$ for $v \in L^1$. Then*

$$(E^{\mathcal{G}}h)(x,\omega) = x \cdot (E^{\mathcal{G}}v)(\omega).$$

Theorem 4.7. *Let h be a Caratheodory function such that there exist $x \in \mathbb{R}^n$ and $\rho \in L^1$ with $h(x) \in L^1$ and*

$$|h(x) - h(x')| \leq \rho|x - x'| \quad \forall x, x' \in \mathbb{R}^n.$$

Then $E^{\mathcal{G}}h$ exists, is unique and characterized by

$$(E^{\mathcal{G}}h)(x) = E^{\mathcal{G}}[h(x)] \quad \forall x \in \mathbb{R}^n.$$

Moreover,

$$|(E^{\mathcal{G}}h)(x - x')| \leq (E^{\mathcal{G}}\rho)|x - x'| \quad \forall x, x' \in \mathbb{R}^d.$$

Proof. The assumptions imply that $h(x) \in L^1$ for all $x \in \mathbb{R}^d$. Let D be a countable dense set in $\mathbb{R}^d$ and, for each $x \in D$, let $\tilde{h}(x)$ be the conditional expectation of $h(x)$. There is a P-null set N such that, outside N,

$$|\tilde{h}(x) - \tilde{h}(x')| \leq (E^{\mathcal{G}}\rho)|x - x'|$$

for each $x, x' \in D$. Outside N, $\tilde{h}$ extends to whole $\mathbb{R}^d$ by continuity while on N we set $\tilde{h} = 0$.

For a simple $x = \sum_{n=1}^N x^n \mathbb{1}_{A^n}$, where $A^n \in \mathcal{G}$ form a disjoint partition of Ω and $x^n \in D$, we have

$$E^{\mathcal{G}}[h(x)] = E^{\mathcal{G}}[h(\sum_{n=1}^N x^n \mathbb{1}_{A^n})] = E^{\mathcal{G}}[\sum_{n=1}^N \mathbb{1}_{A^n} h(x^n)] = \sum_{n=1}^N \mathbb{1}_{A^n}\tilde{h}(x^n) = \tilde{h}(x)$$

Any $x \in L^\infty(\mathcal{G})$ is a pointwise limit of such simple random variables bounded by $\|x\|$, so dominated convergence for conditional expectation and scenariowise continuity of h and $\tilde{h}$ imply that $E^{\mathcal{G}}[h(x)] = \tilde{h}(x)$. Thus Lemma 4.6 implies the claim. □

Lemma 4.8. *Let h^1 and h^2 be L-bounded normal integrands with $h^1 \leq h^2$. Then $E^{\mathcal{G}} h^1 \leq E^{\mathcal{G}} h^2$ whenever the conditional expectations exist.*

Proof. For any $x \in L^\infty(\mathcal{G})$, $h^1(x)$ and $h^2(x)$ are quasi-integrable, so $h^1 \leq h^2$ implies $E^{\mathcal{G}} h^1(x) \leq E^{\mathcal{G}} h^2(x)$ and the result follows from Theorem 3.28. □

The following result is a monotone convergence theorem for conditional expectations of integrands.

Theorem 4.9. *Let $(h^\nu)_{\nu=1}^\infty$ be a nondecreasing sequence of L-bounded normal integrands and*

$$h = \sup_\nu h^\nu.$$

If each $E^{\mathcal{G}} h^\nu$ exists, then $E^{\mathcal{G}} h$ exists and

$$E^{\mathcal{G}} h = \sup_\nu E^{\mathcal{G}} h^\nu.$$

Proof. For any $x \in L^\infty(\mathcal{G})$ and $\alpha \in L^\infty(\mathcal{G}; \mathbb{R}_+)$, monotone convergence and Lemma 4.8 imply that

$$\begin{aligned} E[\alpha h(x)] &= E[\alpha \sup_\nu h^\nu(x)] = \sup_\nu E[\alpha h^\nu(x)] \\ &= \sup_\nu E[\alpha E^{\mathcal{G}}[h^\nu(x)]] = E[\alpha \sup_\nu E^{\mathcal{G}}[h^\nu(x)]]. \end{aligned}$$

Thus $\sup_\nu E^{\mathcal{G}} h^\nu = E^{\mathcal{G}} h$. □

The following gives existence and uniqueness of the conditional expectation of general L-bounded normal integrand.

Theorem 4.10. *An L-bounded normal integrand has a unique conditional expectation and that is L-bounded as well.*

Proof. Let h be an L-bounded normal integrand. Assume first that $h \leq n$ for some constant $n > 0$. By Corollary 3.13,

$$h^\nu(x, \omega) := \inf_{x'} \{h(x', \omega) + \nu \rho(\omega)|x - x'|\}$$

form a nondecreasing sequence of Caratheodory functions increasing pointwise to h. By Theorems 4.7 and 4.9, $E^{\mathcal{G}} h$ exists. To remove the assumption $h \leq n$, consider the nondecreasing sequence $h^n(x) := \min\{h(x), n\}$ and apply Theorem 4.9 again. □

Corollary 4.11. *Assume that S is a closed random set. Then*

$$E^{\mathcal{G}}\delta_S = \delta_{^{\mathcal{G}}S},$$

where $^{\mathcal{G}}S$ is the unique $\mathcal{G}$-measurable closed random set such that $L^0(\mathcal{G}; {}^{\mathcal{G}}S) = L^0(\mathcal{G}; S)$.

Proof. Exercise. Hint: Show first that the conditional expectation of a $\{0, +\infty\}$-valued random variable is $\{0, +\infty\}$-valued. □

The random set $^{\mathcal{G}}S$ in Corollary 4.11 is known as the *$\mathcal{G}$-measurable core of S.*

Lemma 4.12. *Assume that h is a nonnegative normal integrand such that*

$$\inf_{x \in L^0(\mathcal{G})} Eh(x) = 0.$$

Then

$$^{\mathcal{G}}(\operatorname{argmin} h) = {}^{\mathcal{G}}(\operatorname{argmin} E^{\mathcal{G}}h).$$

Proof. Exercise. □

4.1.2 Tower property and independence

Theorem 4.13 (Tower property)**.** *Assume that h is an L-bounded normal integrand, and $\tilde{\mathcal{G}}$ is a sub-σ-algebra of $\mathcal{G}$. Then*

$$E^{\tilde{\mathcal{G}}}(E^{\mathcal{G}}h) = E^{\tilde{\mathcal{G}}}h.$$

Proof. By Theorem 4.10, $E^{\mathcal{G}}h$ exists and is L-bounded, so the result follows from the usual tower property of conditional expectation (see appendix??) and Lemma 4.6. □

Given $\mathcal{H} \subset \mathcal{F}$, σ-algebras $\mathcal{G}'$ and $\mathcal{G}$ are *$\mathcal{H}$-conditionally independent* if

$$E^{\mathcal{H}}[1_{A'}1_A] = E^{\mathcal{H}}[1_{A'}]E^{\mathcal{H}}[1_A]$$

for every $A' \in \mathcal{G}'$ and $A \in \mathcal{G}$.

Lemma 4.14. *Given σ-algebras $\mathcal{G}'$, $\mathcal{G}$ and $\mathcal{H}$, the following are equivalent:*

1. *$\mathcal{G}'$ and $\mathcal{G}$ are $\mathcal{H}$-conditionally independent,*
2. *$E^{\mathcal{H}}[w'w] = E^{\mathcal{H}}[w']E^{\mathcal{H}}[w]$ for every $w' \in L^1(\mathcal{G}')$ and $w \in L^\infty(\mathcal{G})$,*
3. *$E^{\mathcal{G}\vee\mathcal{H}}[w'] = E^{\mathcal{H}}[w']$ for every $w' \in L^1(\mathcal{G}')$.*

Proof. The first implies the second by the monotone class theorem. When 2 holds, we have, for any $w' \in L^1(\mathcal{G}')$, $A \in \mathcal{G}$ and $B \in \mathcal{H}$,

$$E[E^{\mathcal{H}}[w']1_{A\cap B}] = E[E^{\mathcal{H}}[w'1_A]1_B] = E[w'1_A1_B] = E[E^{\mathcal{G}\vee\mathcal{H}}[w']1_{A\cap B}],$$

and, by the monotone class theorem, this extends from sets of the form $A \cap B$ to any set in $\mathcal{G} \vee \mathcal{H}$. Thus 2 implies 3. Assuming 3, we have, for $A' \in \mathcal{G}'$ and $A \in \mathcal{G}$,

$$\begin{aligned} E^{\mathcal{H}}[1_A 1_{A'}] &= E^{\mathcal{H}}[E^{\mathcal{G}\vee\mathcal{H}}[1_A 1_{A'}]] = E^{\mathcal{H}}[1_A E^{\mathcal{G}\vee\mathcal{H}} 1_{A'}] \\ &= E^{\mathcal{H}}[1_A E^{\mathcal{H}} 1_{A'}] = E^{\mathcal{H}}[1_A] E^{\mathcal{H}}[1_{A'}], \end{aligned}$$

so 1 holds. □

A random variable w is $\mathcal{H}$-*conditionally independent* of $\mathcal{G}$ if $\sigma(w)$ is so. Likewise, we say that a normal integrand h is $\mathcal{H}$-*conditionally independent* of $\mathcal{G}$ if $\sigma(h)$ is $\mathcal{H}$-conditionally independent of $\mathcal{G}$. Here $\sigma(h)$ is the smallest σ-algebra under which $\operatorname{epi} h$ is measurable.

Theorem 4.15. *Let h be an L-bounded normal integrand $\mathcal{H}$-conditionally independent of $\mathcal{G}$. Then $E^{\mathcal{G}\vee\mathcal{H}}h = E^{\mathcal{H}}h$. In particular, if h is independent of $\mathcal{G}$, then $E^{\mathcal{G}}h$ is deterministic.*

Proof. Assume first that h satisfies the assumptions of Theorem 4.7. Then $E^{\mathcal{H}}h$ is characterized by

$$E^{\mathcal{H}}(h(x)) = (E^{\mathcal{H}}h)(x) \quad \forall x \in \mathbb{R}^d$$

and likewise for $E^{\mathcal{G}\vee\mathcal{H}}h$. Thus $E^{\mathcal{G}\vee\mathcal{H}}h = E^{\mathcal{H}}h$, by Lemma 4.14. The first claim now follows using Lipschitz regularizations as in the proof of Theorem 4.10. The second claim follows by taking $\mathcal{H}$ the trivial σ-algebra. □

Exercise 4.1.2. *Give two examples of random variables and σ-algebras such that*

1. *the random variables are independent but not conditionally independent.*
2. *the random variables are conditionally independent but not independent.*

Lemma 4.16. *Assume that h is an L-bounded normal integrand and that $\mathcal{G} \neq \mathcal{F}$. If $h(x)$ is $\mathcal{G}$-measurable for $x \in L^\infty$, then h is independent of x, i.e., $h(x,\omega) = h(x',\omega)$ for all x outside a null set.*

Proof. This lemma is false; see the exercises today. □

4.1.3 Convexity

Theorem 4.17. *For a convex L-bounded normal integrand h, $E^{\mathcal{G}}h$ is a convex L-bounded normal integrand. If h is positively homogeneous, so is $E^{\mathcal{G}}h$.*

Proof. Let $\alpha^1, \alpha^2 \geq 0$ and

$$\tilde{h}(x^1, x^2, \omega) := h(\alpha^1 x^1 + \alpha^2 x^2, \omega),$$
$$\bar{h}(x^1, x^2, \omega) := \alpha^1 h(x^1, \omega) + \alpha^2 h_t(x^2, \omega).$$

Both are normal integrands by Theorem 3.25. By Theorem 4.26,

$$E^{\mathcal{G}}\tilde{h}(x^1, x^2, \omega) = E^{\mathcal{G}}h(\alpha^1 x^1 + \alpha^2 x^2, \omega),$$
$$E^{\mathcal{G}}\bar{h}(x^1, x^2, \omega) = \alpha^1 E^{\mathcal{G}}h(x^1, \omega) + \alpha^2 E^{\mathcal{G}}h(x^2, \omega).$$

Since h is convex, $\tilde{h} \leq \bar{h}$ when $\alpha^1 + \alpha^2 = 1$, so Lemma 4.8 implies that $E^{\mathcal{G}}\tilde{h} \leq E^{\mathcal{G}}\bar{h}$. Thus $E^{\mathcal{G}}h$ is convex as well. Positive homogeneity follows similarly by considering any $\alpha^1, \alpha^2 \geq 0$. □

Theorem 4.18. *Let $h : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ be such that, for almost every ω, the function $h(\cdot, \omega)$ is convex and its domain has nonempty interior. Then $E^{\mathcal{G}}h$ exists, is unique and convex and characterized by*

$$(E^{\mathcal{G}}h)(x) = E^{\mathcal{G}}[h(x)] \quad \forall x \in \mathbb{R}^n.$$

Proof. ?? □

Theorem 4.19. *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 $\operatorname{dom} h$ is $\mathcal{G}$-measurable closed-mapping. Then $E^{\mathcal{G}}h$ exists is unique and convex and characterized by*

$$(E^{\mathcal{G}}f)(x) = E^{\mathcal{G}}[f(x)] \quad \forall x \in \mathbb{R}^n.$$

Proof. ?? □

A $\mathcal{G}$-*conditional expectation* of an $\mathcal{F}$-measurable set-valued mapping $S : \Omega \rightrightarrows \mathbb{R}^n$ is a $\mathcal{G}$-measurable closed-valued mapping $E^{\mathcal{G}}S$ such that

$$L^1(\mathcal{G}; E^{\mathcal{G}}S) = \operatorname{cl}\{E^{\mathcal{G}}v \mid v \in L^1(\mathcal{F}; S)\},$$

where the closure is in L^1. The following lemma gives the existence and uniqueness. It is possible to show that $E^{\mathcal{G}}S$ is the smallest $\mathcal{G}$-measurable closed-valued mapping with the property that $E^{\mathcal{G}}v, v \in L^1(\mathcal{F}, S)$ is its selection.

Theorem 4.20. *Assume that S is closed convex-valued mapping and admits at least one integrable selection. Then $E^{\mathcal{G}}S$ exists, is unique, closed-convex valued and characterized by*

$$\delta_{E^{\mathcal{G}}S} = (E^{\mathcal{G}}\sigma_S)^*.$$

Proof. By Theorem 4.17, $E^{\mathcal{G}}\sigma_S$ is a positively homogeneous convex normal integrand. By Theorems 3.26 and **??**, its conjugate is the indicator function of some $\mathcal{G}$-measurable closed convex-valued Γ.

Let $C = \{E^{\mathcal{G}}v \mid v \in L^1(\mathcal{F}; S)\}$. For every $x \in L^\infty(\mathcal{G})$,

$$\sigma_C(x) = E\sigma_S(x) = E[E^{\mathcal{G}}\sigma_S(x)] = E[\sigma_\Gamma(x)] = \sigma_{L^1(\Gamma)}(x),$$

where the first and the last equality follow from Theorem 3.29 and the second from quasi-integrability of $\sigma_S(x)$. By the biconjugate theorem (see appendix), $\operatorname{cl} C = L^1(\Gamma)$, so $\Gamma = E^{\mathcal{G}}S$ by definition. □

Given a convex normal integrand h, the mapping $E^{\mathcal{G}} \operatorname{epi} h$ is an epigraph of a $\mathcal{G}$-normal integrand. We denote this $\mathcal{G}$-normal integrand by ${}^{\mathcal{G}}h$ and call it the *epi-conditional expectation* of h.

Theorem 4.21. *Assume that h is a convex normal integrand such that there exists $x \in \operatorname{dom} Eh \cap L^0(\mathcal{G})$ and $v \in \operatorname{dom} Eh^* \cap L^1$ with $(x \cdot v)^- \in L^1$. Then*

$$(E^{\mathcal{G}}h)^* = {}^{\mathcal{G}}(h^*).$$

Proof. By assumption, there exists $v \in L^1$ such that $(v, h^*(v)^+)$ is an integrable selection of $\operatorname{epi} h^*$. Applying Theorem 4.20 to $\operatorname{epi} h^*$, we get that $E^{\mathcal{G}} \operatorname{epi} h^*$ exists and is characterized by

$$\delta_{E^{\mathcal{G}} \operatorname{epi} h^*} = (E^{\mathcal{G}}\sigma_{\operatorname{epi} h^*})^*.$$

By Lemma 4.29,

$$(E^{\mathcal{G}}\sigma_{\operatorname{epi} h^*})^* = (E^{\mathcal{G}}H)^* = (\sigma_{\operatorname{epi}(E^{\mathcal{G}}h)^*})^* = \delta_{\operatorname{epi}(E^{\mathcal{G}}h)^*}$$

so $\operatorname{epi}(E^{\mathcal{G}}h)^* = E^{\mathcal{G}} \operatorname{epi} h^* = \operatorname{epi}({}^{\mathcal{G}}h^*)$. □

Theorem 4.22 (Jensen's inequality)**.** *If the conjugate of h satisfies the assumptions of Theorem* 4.21*, then*

$$E^{\mathcal{G}}[h(y)] \geq {}^{\mathcal{G}}h(E^{\mathcal{G}}y)$$

for all $y \in L^1(\mathcal{F})$.

Proof. Given integrable $(y, h(y)) \in \operatorname{epi} h$, we have $(E^{\mathcal{G}}y, E^{\mathcal{G}}(h(y))) \in \operatorname{epi} {}^{\mathcal{G}}h$. □

Lemma 4.23. *A random closed convex set S with $L^1(S) \neq \emptyset$ is $\mathcal{G}$-measurable if and only if*

$$E^{\mathcal{G}}S \subseteq S$$

almost surely.

Proof. It suffices to prove that $\omega \mapsto d(x, S(\omega))$ is $\mathcal{G}$-measurable for every $x \in \mathbb{R}^n$. By the measurable selection theorem, there exists $y \in L^0(S)$ such that

$$d(x, S(\omega)) = |x - y(\omega)|.$$

Since $L^1(S) \neq \emptyset$, $d(x, S)$ is integrable, so $y \in L^1(S)$. By Jensen's inequality,

$$Ed(x, S) \geq E|x - E^{\mathcal{G}} y|.$$

Since $E^{\mathcal{G}} y \in S$, we also have $|x - E^{\mathcal{G}} y| \geq d(x, S)$, so $d(x, S) = |x - E^{\mathcal{G}} y|$. □

Corollary 4.24. *If the conjugate of h satisfies the assumptions of Theorem 4.21, then h is $\mathcal{G}$-measurable if and only if*

$$E^{\mathcal{G}}[h(y)] \geq h(E^{\mathcal{G}} y)$$

for all $y \in L^1(\mathcal{F})$.

Proof. Necessity follows from Theorem 4.22. By Lemma 4.23, it suffices to show that $E^{\mathcal{G}} \operatorname{epi} h \subset \operatorname{epi} h$. Any $(y, \alpha) \in E^{\mathcal{G}} \operatorname{epi} h$ is a L^1-limit of $(E^{\mathcal{G}} y^\nu, \alpha^\nu)$ for integrable $(y^\nu, \alpha^\nu) \in \operatorname{epi} h$. By assumption,

$$E^{\mathcal{G}} \alpha^\nu \geq E^{\mathcal{G}} h(y^\nu) \geq h(E^{\mathcal{G}} y^\nu),$$

so $(E^{\mathcal{G}} y^\nu, E^{\mathcal{G}} \alpha^\nu) \in \operatorname{epi} h$, which proves the claim. □

The following is a counterexample to some of the questions posed in the lectures. Given an integrable process x,

$$x \mapsto {}^a x := (E_t x_t)_{t=0}^T.$$

is the *adapted projection* of x.

Exercise 4.1.3. *Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, $\mathcal{G} = \sigma([0, a] \mid a \leq 1/2)$, $B = [0, 1/2]$ and*

$$h(x, \omega) := 1_B |x| = 1_B (|x_0|^2 + |x_1|^2)^{1/2}.$$

Let $\mathcal{F}_1 := \mathcal{F}$, $\mathcal{F}_0 := \mathcal{G}$. Show that

1. *the composition $h(x)$ is $\mathcal{G}$-measurable even though h depends on x.*
2. *The "Jensen's inequality for processes" $h(\bar{x}) \geq h({}^a \bar{x})$ holds even though h is not of the form $\sum_{t=0}^T h_t(x_t, \omega)$.*

4.1.4 Algebraic operations

Lemma 4.25. *Let $\xi_1, \xi_2 \in L^0(\mathcal{F};\overline{\mathbb{R}})$ be quasi-integrable. Then*

1. *$E^{\mathcal{G}}[\xi_1+\xi_2] = E^{\mathcal{G}}[\xi_1] + E^{\mathcal{G}}[\xi_2]$ if ξ_1, ξ_2 satisfy any of the conditions in Lemma 3.33.*
2. *$E^{\mathcal{G}}[\xi_1\xi_2] = \xi_1 E^{\mathcal{G}}[\xi_2]$ if $\xi_1 \in L^\infty(\mathcal{G})$.*

Proof. Let $\alpha \in L^\infty_+(\mathcal{G})$. To prove 1, Lemma 3.33 gives

$$E[\alpha(\xi_1+\xi_2)] = E[\alpha\xi_1]+E[\alpha\xi_2] = E[\alpha E^{\mathcal{G}}\xi_1]+E[\alpha E^{\mathcal{G}}\xi_2] = E[\alpha(E^{\mathcal{G}}\xi_1+E^{\mathcal{G}}\xi_2)].$$

The second claim is clear. □

Theorem 4.26 (Conditional expectation in operations)**.** *Let h, h^1 and h^2 be L-bounded normal integrands.*

1. *h^1+h^2 is L-bounded and $E^{\mathcal{G}}(h^1+h^2) = E^{\mathcal{G}}h^1 + E^{\mathcal{G}}h^2$.*
2. *If $\alpha \in L^\infty_+(\mathcal{G})$, then αh is L-bounded and $E^{\mathcal{G}}(\alpha h) = \alpha E^{\mathcal{G}} h$.*
3. *If $M : \Omega\times\mathbb{R}^n \to \mathbb{R}^d$ is a $\mathcal{G}$-measurable Caratheodory mapping such that $|M(x)| \le \alpha|x| + b$ for $\alpha, \beta \in L^0$ with $\rho\alpha$ and $\rho\beta$ integrable, then $h^1\circ M$ is L-bounded and $E^{\mathcal{G}}(h\circ M) = (E^{\mathcal{G}}h)\circ M$.*

Proof. For any $x \in L^\infty(\mathcal{G})$, $h^1(x)^-$ and $h^2(x)^-$ are integrable, so 1 follows from Lemmas 4.6 and 4.25. The second follows similarly from Lemmas 4.6 and 4.25. To prove 3, assumptions imply that $h\circ M$ satisfies the assumption of Theorem 4.10 and that $h(M(x)$ is quasi-integrable for every $x \in L^\infty$. Thus Theorem 4.10 implies $E^{\mathcal{G}}(h(M(x)) = (E^{\mathcal{G}}h)(M(x))$. □

Theorem 4.27. *Let h be a convex normal integrand and F a K-convex $\mathcal{G}$-normal function such that (3.1) holds and $(-K)\cap\{u \mid h^\infty(u,\omega) \le 0\}$ is linear almost surely, and there exists $y \in \operatorname{dom} Eh^* \cap L^1$ such that $y\cdot F$ is L-bounded. Then $h\circ F$ is an L-bounded convex normal integrand and*

$$E^{\mathcal{G}}(h\circ F) = (E^{\mathcal{G}}h)\circ F.$$

Proof. By Fenchel's inequality,

$$h(F(x,\omega),\omega) \ge y(\omega)\cdot F(x,\omega) - h^*(y),$$

so the composition is L-bounded and the claim follows from Theorem 4.10. □

Theorem 4.28. *Assume that h is a convex normal integrand such that there exists $x \in \operatorname{dom} Eh \cap L^0(\mathcal{G})$ and $v \in \operatorname{dom} Eh^* \cap L^1$ with $(x\cdot v)^- \in L^1$. Then*

$$E^{\mathcal{G}}(h^\infty) = (E^{\mathcal{G}}h)^\infty.$$

Proof. The difference quotients

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

define a nondecreasing sequence of integrands $(h^\lambda)_{\lambda=1}$ with pointwise limit h^∞. Fenchel's inequality gives

$$h^\lambda(x',\omega) \geq x'\cdot v(\omega) + x\cdot v(\omega) - h^*(v(\omega),\omega)) - h(x(\omega),\omega),$$

so Theorems 4.9 and 4.26 imply the claim. ☐

Lemma 4.29. *Assume that h is a convex normal integrand such that there exists $x \in \operatorname{dom} Eh \cap L^0(\mathcal{G})$ and $v \in \operatorname{dom} Eh^* \cap L^1$ with $(x\cdot v)^- \in L^1$. Then*

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

is an L-bounded convex normal integrand with $H^* = \delta_{\operatorname{epi} h^*}$ and

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

Proof. Since $v \in \operatorname{dom} Eh^* \cap L^1$, H is L-bounded.

Assume for a contradiction that there exists $(x,\alpha) \in L^0(\mathcal{G})_+ \times L^0(\mathcal{G})$ such that $E^{\mathcal{G}}\sigma_{\operatorname{epi} h^*}(x,\alpha) \neq \sigma_{\operatorname{epi}(E^{\mathcal{G}}h)^*}(\alpha,x)$. Let

$$A_\nu := \{\omega \mid \alpha(\omega) \in \{0\} \cup [1/\nu,\nu], |x(\omega)| \leq \nu\}.$$

Now L-boundedness of h gives

$$E^{\mathcal{G}}[\sigma_{\operatorname{epi} h^*}(x,\alpha)] = \sigma_{\operatorname{epi}(E^{\mathcal{G}}h)^*}(\alpha,x) \quad \text{on } A_\nu.$$

Since $P(A_\nu) \nearrow 1$, this leads to a contradiction. ☐

4.1.5 Regular conditional distributions

Given a sigma algebra $\mathcal{G} \subseteq \mathcal{F}$ and a measurable space $(\Xi,\mathcal{A})$, a function $\mu : \Omega \times \mathcal{A} \to \overline{\mathbb{R}}$ is a *probability kernel* from $(\Omega,\mathcal{G})$ to $(\Xi,\mathcal{A})$ if for each $\omega \in \Omega$ and $A \in \mathcal{A}$, $\mu(\cdot,A)$ is a $\mathcal{G}$-measurable function on Ω and $\mu(\omega,\cdot)$ is a probability measure on $(\Xi,\mathcal{A})$. Given an $\mathcal{A}$-measurable real-valued function X on $(\Xi,\mathcal{A})$, we will use the notation

$$(E^\mu X)(\omega) := \int_S X(s)\mu(\omega,ds).$$

As soon as $E^\mu X^-$ is real-valued, $E^\mu X$ is $\mathcal{G}$-measurable. More generally, we have the following.

Theorem 4.30. *Let $H : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ be a $\mathcal{A}$-normal integrand and μ be a probability kernel from $(\Omega, \mathcal{G})$ to $(\Xi, \mathcal{A})$. If there exist $R, M : \Xi \to \mathbb{R}_+$ with $E^\mu M$ and $E^\mu R$ real-valued and*

$$H(x,s) \geq -R(s)|x| - M(s) \quad \mu(\omega,\cdot)\text{-a.e.}$$

for all $\omega \in \Omega$, then

$$(E^\mu H)(x,\omega) := \int_\Xi H(x,s)\mu(\omega,ds)$$

is a $\mathcal{G}$-normal integrand with

$$E^\mu H(x,\omega) \geq -(E^\mu R)(\omega)|x| - (E^\mu M)(\omega).$$

Proof. The function $E^\mu H$ is $\mathcal{B}(\mathbb{R}^n)\otimes\mathcal{F}$-measurable. Indeed, this clearly holds if $H = 1_B 1_A$ for $B \in \mathcal{B}(\mathbb{R}^n)$ and $A \in \mathcal{A}$, and a monotone class argument extends this to every $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{A}$-measurable nonnegative H. Recalling the lower bound for H, the general case follows by taking differences.

Let

$$H^\nu(x,s) := \inf_{x'}\{H(x',s) + \nu R(s)|x - x'|\},$$
$$(E^\mu H^\nu)(x,\omega) := \int_S H^\nu(x,s)\mu(\omega,ds).$$

By Lemma 8.12 in the appendix, $\mu(\omega,\cdot)$-a.e., each $H^\nu(\cdot,s)$ is $(\nu R(s))$-Lipschitz, $H^\nu(x,s)$ increase pointwise to $H(x,s)$ and

$$H^\nu(x,s) \geq -R(s)|x| - M(s).$$

By Jensen's inequality and the Lipschitz property,

$$\begin{aligned}
|(E^\mu H^\nu)(x,\omega) - (E^\mu H^\nu)(x',\omega)| &\leq \int_S |H^\nu(x,s) - H^\nu(x',s)|\mu(\omega,ds) \\
&\leq \int_S \nu R(s)|x - x'|\mu(\omega,ds) \\
&\leq \nu(E^\mu R)(\omega)|x - x'|.
\end{aligned}$$

Thus, since $E^\mu H^\nu$ is measurable, it is a Caratheodory integrand and hence normal. Since H^ν have a common lower bound, monotone convergence gives

$$(E^\mu H)(x,\omega) = \sup_\nu (E^\mu H^\nu)(x,\omega),$$

so $E^\mu H$ is a normal integrand by Theorem 4.9. □

A probability kernel μ from $(\Omega,\mathcal{G})$ to $(\Xi,\mathcal{A})$ is a *regular $\mathcal{G}$-conditional distribution* of an $\eta \in L^0(\mathcal{F};\Xi)$ if

$$E^\mathcal{G}[X(\eta)] = (E^\mu X)$$

almost surely for every nonnegative $\mathcal{A}$-measurable function X on Ξ.

Lemma 4.31. *Given an $\eta \in L^0(\mathcal{F};\Xi)$, a $\sigma(\eta)$-measurable normal integrand h is L-bounded if and only if*

$$h(x,\omega) = H(x,\eta(\omega)),$$

where $H : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ is an $\mathcal{A}$-normal integrand such that there exist $\mathcal{A}$-measurable $M, R : \Xi \to \mathbb{R}_+$ such that $R(\eta), M(\eta) \in L^1$ and

$$H(x,\eta(\omega)) \geq -R(\eta(\omega))|x| - M(\eta(\omega)) \quad \forall x \in \mathbb{R}^n$$

almost surely. If η has a regular $\mathcal{G}$-conditional distribution μ, then these equivalent conditions imply

$$H(x,s) \geq -R(s)|x| - M(s) \quad \forall x \in \mathbb{R}^n \;\; \mu(\omega,\cdot)\text{-a.e.}$$

for P-almost every ω.

Proof. It is clear that the stated conditions imply L-boundedness of h. On the other hand, $\sigma(\eta)$-measurability of h implies, by Corollary 3.17, that there exists an $\mathcal{A}$-normal integrand $H : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ such that $h(x,\omega) = H(x,\eta(\omega))$. By L-boundedness, there exists $\rho, m \in L^1$ such that

$$h(x,\omega) \geq -\rho(\omega)|x| - m(\omega) \quad \forall x \in \mathbb{R}^n.$$

Since h is $\sigma(\eta)$-measurable, $E^{\sigma(\eta)}h = h$ and

$$h(x,\omega) \geq -(E^{\sigma(\eta)}\rho)(\omega)|x| - (E^{\sigma(\eta)}m)(\omega) \quad \forall x \in \mathbb{R}^n$$

almost surely. By the Doob-Dynkin lemma for normal integrands (Corollary 3.17), there exist measurable $R, M : \Xi \to \mathbb{R}$ such that $(E^{\sigma(\eta)}\rho)(\omega) = R(\eta(\omega))$ and $(E^{\sigma(\eta)}m)(\omega) = M(\eta(\omega))$.

Assume now that the above holds and let

$$X(s) = \inf_{x\in\mathbb{R}^n}\{H(x,s) + R(s)|x| + M(s)\}.$$

By Corollary 3.11, X is $\mathcal{A}$-measurable. If $X(\eta) \geq 0$ almost surely, then $1_{\{X<0\}}(\eta) = 0$ almost surely so

$$E^{\mu}1_{\{X<0\}} = E^{\mathcal{G}}[1_{\{X<0\}}(\eta)] = 0$$

almost surely. In other words,

$$X(s) \geq 0 \quad \mu(\omega,\cdot)\text{-a.e.}$$

almost surely. □

Theorem 4.32 (Disintegration)**.** *Let*

$$h(x,\omega) := H(x,\eta(\omega))$$

where $\eta \in L^0(\mathcal{F};\Xi)$ and $H : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ is an $\mathcal{A}$-normal integrand. If h is L-bounded and μ is a regular $\mathcal{G}$-conditional distribution of η, then

$$E^{\mathcal{G}}h = E^{\mu}H$$

almost surely everywhere.

Proof. By Lemma 4.31, there exist $\mathcal{A}$-measurable $M, R : \Xi \to \mathbb{R}_+$ such that $R(\eta), M(\eta) \in L^1$ and

$$H(x, \eta(\omega)) \geq -R(\eta(\omega))|x| - M(\eta(\omega)) \quad \forall x \in \mathbb{R}^n$$

almost surely, and

$$H(x, s) \geq -R(s)|x| - M(s) \quad \forall x \in \mathbb{R}^n \ \mu(\omega, \cdot)\text{-a.e.} \tag{4.3}$$

for P-almost every ω. Let $N \in \mathcal{F}$ be a P-null set such that $E^\mu R$ and $E^\mu M$ are real-valued and (4.3) holds on N^C. By Theorem 4.30, $E^\mu H$ is a $\mathcal{G}$-normal integrand relative to N^C so, by Lemma 4.6, it suffices to show that

$$E^\mu H(x) = E^{\mathcal{G}}[H(x, \eta)]$$

almost surely for all $x \in L^\infty(\mathcal{G}; \mathbb{R}^n)$. This clearly holds if $H = 1_B 1_A$ for $B \in \mathcal{B}(\mathbb{R}^n)$ and $A \in \mathcal{A}$, and a monotone class argument extends this to every $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{A}$-measurable nonnegative H. Recalling the lower bound for H, the general case follows by taking differences?? □

The following is a special of case Theorem 4.32 with $(\Xi, \mathcal{A}) = (\Omega, \mathcal{F})$ and $\eta(\omega) = \omega$.

Corollary 4.33. *Let h be an L-bounded normal integrand. If the identity mapping on Ω has a regular $\mathcal{G}$-conditional distribution μ, then*

$$E^{\mathcal{G}} h = E^\mu h.$$

Given $\xi \in L^0(\mathbb{R}^d)$, a probability kernel ν from $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ to $(\Xi, \mathcal{A})$ is called a *regular ξ-conditional distribution of $\eta \in L^0(\Xi)$* if $\mu(\omega, B) := \nu(\xi(\omega), B)$ defines a $\sigma(\xi)$-conditional regular distribution of η. If X is an $\mathcal{A}$-measurable function on Ξ such that $X(\eta(\cdot))$ is quasi-integrable, then, almost surely,

$$E^{\sigma(\xi)}[X(\eta)](\omega) = (E^\nu X)(\xi(\omega)).$$

More generally, applying Lemma 4.31 and Theorem 4.32 with $\mu(\omega, B) := \nu(\xi(\omega), B)$ gives the following.

Corollary 4.34. *Let*

$$h(x, \omega) := H(x, \eta(\omega))$$

where $\eta \in L^0(\mathcal{F}; \Xi)$ and $H : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ is a $\mathcal{A}$-normal integrand. If h is L-bounded and ν is a regular ξ-conditional distribution of η, then

$$(E^{\sigma(\xi)} h)(x, \omega) = (E^\nu H)(x, \xi(\omega)).$$

Corollary 4.35. *Let*

$$h(x, \omega) := H(x, \xi_1(\omega), \xi_2(\omega))$$

where $\xi_1, \xi_2 \in L^0(\mathcal{F};\mathbb{R}^m)$ *and* $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$ *is a* $\mathcal{B}(\mathbb{R}^m \times \mathbb{R}^m)$*-normal integrand. If* h *is* L*-bounded and* ν *is a regular* ξ_1*-conditional distribution of* ξ_2*, then*

$$(E^{\sigma(\xi_1)}h)(x,\omega) = (E^\nu H)(x,\xi_1(\omega)),$$

where

$$(E^\nu H)(x,\xi_1) := \int_\Xi H(x,\xi_1,s)\nu(\xi_1,ds).$$

If, in addition, $\sigma(\xi_1) \subseteq \mathcal{G}$ *and* $\mathcal{F}$ *is* $\sigma(\xi_1)$*-conditionally independent of* $\mathcal{G}$*, then*

$$(E^\mathcal{G} h)(x,\omega) = (E^\nu H)(x,\xi_1(\omega)).$$

Proof. The first claim follows from Corollary 4.34 with $\eta = (\xi_1, \xi_2)$ and $\xi = \xi_1$. The second claim follows from the first and Theorem 4.15. □

Exercise 4.1.4. *In the last setting of the above corollary, assume that* ξ_1, ξ_2 *are real-valued independent random variables. Show that*

$$(E^\mathcal{G} h)(x,\omega) = EH(x,s_1,\xi_2)_{|s_1=\xi_1(\omega)}.$$

Hint: show first that the regular ξ_1*-conditional distribution* ν *of* ξ_2 *is given by*

$$\nu(\xi_1(\omega),B) = \Phi_{\xi_2}(B),$$

where $\Phi_{\xi_2}(B) := P(\xi_2 \in B)$ *defines the distribution of* ξ_2*.*

Assume that the joint distribution of (ξ_1,ξ_2) has a probability density ϕ, i.e.,

$$P((\xi_1,\xi_2) \in B) = \int_B \phi(s_1,s_2)d(s_1,s_2)$$

(integral w.r.t. to the Lebesque measure). Then the regular ξ_1-conditional distribution ν of ξ_2 has the representation (we omit the details of this fact)

$$\nu(s_1,B_2) = \frac{\int_{B_2}\phi(s_1,s_2)ds_2}{\phi_1(s_1)},$$

where $\phi_1(s_1) := \int \phi(s_1,s_2)ds_2$ is the probability density of the distribution of ξ_1. Thus, in the setting of Corollary 4.35,

$$(E^{\sigma(\xi_1)}h)(x,\omega) = \frac{\int H(x,\xi_1(\omega),s_2)\phi(\xi_1(\omega),s_2)ds_2}{\phi_1(\xi_1(\omega))}.$$

If (ξ_1,ξ_2) are independent then $\phi(s_1,s_2) = \phi_1(s_1)\phi_2(s_2)$ and the formula reduces, as in the above exercise, to

$$(E^{\sigma(\xi_1)}h)(x,\omega) = \int H(x,\xi_1(\omega),s_2)\phi_2(s_2)ds_2.$$

4.2 Existence of solutions

Recall that an adapted sequence $(h_t)_{t=0}^T$ of normal integrands $h_t : \mathbb{R}^{n^t} \times \Omega \to \overline{\mathbb{R}}$ solves the generalized Bellman equations for h if

$$\begin{aligned} \tilde{h}_T &:= h, \\ h_t &:= E_t \tilde{h}_t \\ \tilde{h}_{t-1}(x^{t-1}, \omega) &:= \inf_{x_t \in \mathbb{R}^{n_t}} h_t(x^{t-1}, x_t, \omega), \end{aligned} \tag{BE}$$

for $t = T, \dots, 0$. The following result gives sufficient conditions for the existence of a unique solution of (BE).

Theorem 4.36 (Bellman equations)**.** *Assume that h is convex and that*

1. *there exist $p \in \mathcal{N}^\perp$ and $m \in L^1$ such that*

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

2. *if h_t is well-defined by* (BE)*, then the set*

$$N_t(\omega) := \{x_t \in \mathbb{R}^{n_t} \mid h_t^\infty(x^t, \omega) \leq 0,\ x^{t-1} = 0\}$$

is linear.

Then (BE) *has a unique solution $(h_t)_{t=0}^T$ and*

$$h_t(x^t, \omega) \geq x^t \cdot (E_t p^t)(\omega) - (E_t m)(\omega) \tag{4.4}$$

for all $t = 0, \dots, T$.

Proof. Assume that h_{t+1} is a well-defined normal integrand and that

$$h_{t+1}(x^{t+1}, \omega) \geq x^{t+1} \cdot (E_{t+1} p^{t+1})(\omega) - (E_{t+1} m)(\omega).$$

By Theorem 3.24, $\tilde{h}_t$ is a normal integrand. Since $E_{t+1} p_{t+1}^{t+1} = 0$, $\tilde{h}_t$ is L-bounded, so h_t exists and is unique by Theorem 4.10. By the same theorem, the lower bound admits a conditional expectation, so we get

$$h_t(x, \omega) \geq x^t \cdot (E_t p^t)(\omega) - (E_t m)(\omega).$$

By induction, (BE) admits a unique solution satisfying (4.4). □

Corollary 4.37. *In the setting of Theorem 4.36, assume that* (SP_u) *is feasible. Given the solution $(h_t)_{t=0}^T$ of* (BE) *for h, $(h_t^\infty)_{t=0}^T$ is the unique solution of* (BE) *for h^∞.*

Proof. If $x \in \operatorname{dom} Eh \cap \mathcal{N}$ and $(h_t)_{t=0}^T$ solves (BE) for h, then $x^t \in \operatorname{dom} E\tilde{h}_t$. Thus, by Theorem 4.28, $h_t^\infty = E_t(\tilde{h}_t^\infty)$, while, by Theorem 3.24,

$$\tilde{h}_{t-1}^\infty(x_{t-1}, \omega) = \inf_{x_t} h_t^\infty(x_{t-1}, x_t, \omega).$$

By recursion, $(h_t^\infty)_{t=0}^T$ solves (BE) for h^∞. Since h^∞ satisfies the assumptions of Theorem 4.36, the solution is unique. □

Theorem 4.38. *In the setting of Theorem 4.36, assume that*

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

Then

$$\inf(\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} Eh_t(x^t)$$

for all $t = 0, \dots, T$, (SP_u) has a solution $\bar{x} \in \mathcal{N}$ with $\bar{x}_t \perp N_t$ for every $t = 0, \dots, T$ and, moreover, the solutions $\bar{x} \in \mathcal{N}$ of (SP_u) are characterized by

$$\bar{x}_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^{n_t}} h_t(\bar{x}^{t-1}, x_t) \quad a.s. \quad t = 0, \dots, T.$$

Proof. By Theorem 4.36, (BE) has a unique solution $(h_t)_{t=0}^T$. Let

$$k(x, \omega) := h(x, \omega) - x \cdot p(\omega)$$

and

$$k_t(x, \omega) := h_t(x^t, \omega) - x^t \cdot E_t p^t(\omega).$$

Since $E_t p_t^t = 0$, Theorem 4.26 gives, recursively backward in time, that $(k_t)_{t=0}^T$ solves (BE) for k. Since, by assumption, $Eh = Ek$ on $\mathcal{N}$, Theorem 4.2 says that

$$\inf(\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} Ek_t(x^t) \tag{4.5}$$

for all $t = 0, \dots, T$ and that the optimal solutions of (SP_u) are characterized by the condition

$$x_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^{n_t}} k_t(x^{t-1}, x_t) \quad \text{a.s.} \quad t = 0, \dots, T.$$

Since $E_t p_t^t = 0$, we have

$$\operatorname*{argmin}_{x_t} k_t(x^{t-1}(\omega), x_t, \omega) = \operatorname*{argmin}_{x_t} h_t(x^{t-1}(\omega), x_t, \omega)$$

and

$$N_t = \{x_t \in \mathbb{R}^{n_t} \mid k_t^\infty(x^t, \omega) \le 0, \ x^{t-1} = 0\}.$$

Applying Theorem 3.24 recursively forward in time gives an optimal $\bar{x} \in \mathcal{N}$ with $\bar{x}_t \perp N_t$ for all t.

It remains to prove that $\inf(\text{SP}_u) = \inf_{x^t \in \mathcal{N}^t} Eh_t(x^t)$. By the definition of $(h_t)_{t=0}^T$, the inequality $\inf(\text{SP}_u) \geq \inf_{x^t \in \mathcal{N}^t} Eh_t(x^t)$ is clear. If $Eh_t(x^t) < \infty$, then, by (4.4),

$$E[x^t \cdot (E_t p^t)] < \infty,$$

so Lemma 3.39 gives $Ek_t(x^t) = Eh_t(x^t)$. Thus

$$\inf_{x^t \in \mathcal{N}^t} Eh_t(x^t) \geq \inf_{x^t \in \mathcal{N}^t} Ek_t(x^t),$$

so the claim follows from (4.5). □

The following exercise shows that assumptions of Theorem 4.36 are not sufficient for Theorem 4.38.

Exercise 4.2.1. *Let $n_t = 1$, $\alpha \in L^2(\mathcal{F}_0)$ and $p \in \mathcal{N}^\perp$ be such that $E[\alpha p_0] = \infty$ and consider*

$$h(x, \omega) := \frac{1}{2}|x_0 - \alpha(\omega)|^2 + x_0 p_0(\omega).$$

Show that (BE) *has a unique solution, there is a unique x satisfying* (4.9) *but* (SP_u) *does not have a solution.*

The following result shows that the linearity condition of Theorem 4.38 can be stated in terms of the original normal integrand h directly. We will use the consequence of Theorem 3.28 that if C is closed-valued and $\mathcal{G}$-measurable, then it is almost surely linear-valued if and only if the set of its measurable selections is a linear space.

Theorem 4.39. *Assume that* (SP_u) *is feasible and that h is a convex normal integrand and that there exists $p \in \mathcal{N}^\perp$ and $m \in L^1$ such that*

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

and

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

Then h_t are well-defined and

$$N_t = \{x_t \in \mathbb{R}^{n_t} \mid h_t^\infty(x^t, \omega) \leq 0,\ x^{t-1} = 0\}$$

are linear-valued for all t if and only if

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

is a linear space. In this case, if $x \in \mathcal{L}$ is such that $x^{t-1} = 0$ then $x_t \in N_t$ almost surely.

Proof. Let $k(x, \omega) := h(x, \omega) - x \cdot p(\omega)$. As in the proof of Theorem 4.38,

$$N_t = \{x_t \in \mathbb{R}^{n_t} \mid k_t^\infty(x^t, \omega) \leq 0,\ x^{t-1} = 0\}$$

so we may assume that h is bounded from below. The feasibility means that there is an $\tilde{x} \in E \operatorname{dom} h \cap \mathcal{N}$. Making the change of variables, $x \to x - \tilde{x}$, we may assume that Eh is finite at the origin.

Note first that, as soon as h_t is well-defined, Theorem 4.28 gives $(h_t)^\infty = E_t(\tilde{h}^\infty)$ and then, by Lemma 4.12,

$$\{x^t \in L^0(\mathcal{F}_t) \mid \tilde{h}_t^\infty(x^t) \le 0 \ a.s.\} = \{x^t \in L^0(\mathcal{F}_t) \mid h_t^\infty(x^t) \le 0 \ a.s.\}. \qquad (4.6)$$

We proceed by induction on T. When $T = 0$, Theorem 4.10 implies that h_T is well defined so, by (4.6),

$$\mathcal{L} = \{x \in \mathcal{N} \mid h_T^\infty(x) \le 0 \text{ a.s.}\},$$

which, by definition of N_T, equals $L^0(\mathcal{F}_T; N_T)$. Since N_T is $\mathcal{F}_T$-measurable, by Theorem 3.7, the linearity of $\mathcal{L}$ is equivalent to N_T being linear-valued. Let now T be arbitrary and assume that the claim holds for every $(T-1)$-period model. If $\mathcal{L}$ is linear then $\mathcal{L}' := \{x \in \mathcal{N} \mid x_0 = 0,\ h^\infty(x) \le 0 \text{ a.s.}\}$ is linear as well. Applying the induction hypothesis to the $(T-1)$-period model obtained by fixing $x_0 \equiv 0$, we get that h_t is well defined and N_t is linear for $t = T, \dots, 1$. By Theorem 3.24 and Theorem 4.10, h_0 is well defined so, by (4.6),

$$\begin{aligned}
L^0(\mathcal{F}_0; N_0) &= \{x_0 \in L^0(\mathcal{F}_0) \mid h_0^\infty(x_0) \le 0 \text{ a.s.}\}\\
&= \{x_0 \in L^0(\mathcal{F}_0) \mid \tilde{h}_0^\infty(x_0) \le 0 \text{ a.s.}\}\\
&= \{x_0 \in L^0(\mathcal{F}_0) \mid \inf_{x_1 \in \mathbb{R}^{n_1}} h_1^\infty(x_0, x_1) \le 0 \text{ a.s.}\}\\
&= \{x_0 \in L^0(\mathcal{F}_0) \mid \exists \tilde{x} \in \mathcal{N} :\ \tilde{x}_0 = x_0,\ h_1^\infty(\tilde{x}^1) \le 0 \text{ a.s.}\},
\end{aligned}$$

where the last equality follows by applying the last part of Theorem 3.24 to the normal integrand h^∞. Repeating the argument for $t = 1, \dots, T$, we get

$$\begin{aligned}
L^0(\mathcal{F}_0; N_0) &= \{x_0 \in L^0(\mathcal{F}_0) \mid \exists \tilde{x} \in \mathcal{N} : \tilde{x}_0 = x_0,\ h_T^\infty(\tilde{x}) \le 0 \text{ a.s.}\}\\
&= \{x_0 \in L^0(\mathcal{F}_0) \mid \exists \tilde{x} \in \mathcal{N} : \tilde{x}_0 = x_0,\ h^\infty(\tilde{x}) \le 0 \text{ a.s.}\}\\
&= \{x_0 \in L^0(\mathcal{F}_0) \mid \exists \tilde{x} \in \mathcal{L} : \tilde{x}_0 = x_0\}. \qquad (4.7)
\end{aligned}$$

The linearity of $\mathcal{L}$ thus implies that of $L^0(\mathcal{F}_0; N_0)$ which is equivalent to N_0 being linear-valued.

Assume now that N_t is linear-valued for $t = T, \dots, 0$ and let $x \in \mathcal{L}$. Expression (4.7) for $L^0(\mathcal{F}_0; N_0)$ is again valid so, by linearity of N_0, there is an $\tilde{x} \in \mathcal{L}$ with $\tilde{x}_0 = -x_0$. Since h^∞ is sublinear, $\mathcal{L}$ is a cone, so that $x + \tilde{x} \in \mathcal{L}$. Since $x_0 + \tilde{x}_0 = 0$, we also have $x + \tilde{x} \in \mathcal{L}'$. Since, by the induction assumption, $\mathcal{L}'$ is linear and since $\mathcal{L}' \subseteq \mathcal{L}$, we get $-x - \tilde{x} \in \mathcal{L}$. Since $\mathcal{L}$ is a cone, we get $-x = \tilde{x} - x - \tilde{x} \in \mathcal{L}$. Thus, $\mathcal{L}$ is linear.

For $t = 0$, the last claim follows directly from expression (4.7). The general case follows by applying this to the $(T-t)$-period model obtained by fixing $x^{t-1} \equiv 0$. □

For convenience, we summarize the main findings of this section in the following.

Theorem 4.40. *Assume that h is a convex normal integrand such that*

1. *there exists $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ with $\lambda p \in \operatorname{dom} Eh^*$ for all $\lambda \in [1-\epsilon, 1+\epsilon]$,*
2. *$\{x \in \mathcal{N} \,|\, h^\infty(x) - x \cdot p \le 0 \text{ a.s.}\}$ is a linear space.*

Then (BE) *has a unique solution $(h_t)_{t=0}^T$ with*

$$h_t(x^t, \omega) \ge x^t \cdot (E_t p^t)(\omega) - (E_t m)(\omega) \tag{4.8}$$

for all $t = 0, \dots, T$, $(h_t^\infty)_{t=0}^T$ solves (BE) *for h^∞ in place of h and*

$$N_t(\omega) := \{x_t \in \mathbb{R}^{n_t} \mid h_t^\infty(x^t, \omega) \le 0,\ x^{t-1} = 0\}$$

are linear. Moreover, (SP_u) *has a solution $x \in \mathcal{N}$ with $x_t \perp N_t$ for every $t = 0, \dots, T$,*

$$\inf(\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} Eh_t(x^t)$$

for all $t = 0, \dots, T$ and the solutions of (SP_u) *are characterized by*

$$x_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^{n_t}} h_t(x^{t-1}, x_t) \quad a.s. \quad t = 0, \dots, T. \tag{4.9}$$

4.3 Applications

4.3.1 Regular conditional distributions

Example 4.41. *Let $(\xi_t)_{t=0}^T$ be a stochastic process such that each ξ_t takes values in a measurable space $(\Xi_t, \mathcal{A}_t)$. Let $\xi^t := (\xi_0, \dots, \xi_t)$ and $\Xi^t := \prod_{t=0}^t \Xi_t$ be equipped with $\mathcal{A}^t := \mathcal{A}_0 \otimes \cdots \otimes \mathcal{A}_t$. In the setting of Theorem 4.36, assume that $\mathcal{F}_t = \sigma(\xi^t)$ and that*

$$h(x, \omega) = \hat{h}(x, \xi^T(\omega))$$

for a convex normal integrand $\hat{h} : \mathbb{R}^n \times \Xi^T \to \overline{\mathbb{R}}$. If ν_t is a regular ξ^t-conditional distribution of ξ_{t+1}, then the solution $(h_t)_{t=0}^T$ of (BE) *is given by*

$$h_t(x^t, \omega) = \hat{h}_t(x^t, \xi^t(\omega)),$$

where $\hat{h}_t : \mathbb{R}^{n^t} \times \Xi^t \to \overline{\mathbb{R}}$ are convex normal integrands defined recursively by

$$\begin{aligned}
\check{h}_T &:= \hat{h}, \\
\hat{h}_t &:= E^{\nu_t} \check{h}_t, \\
\check{h}_{t-1}(x^{t-1}, s^t) &:= \operatorname{cl}_{x^{t-1}} \inf_{x_t \in \mathbb{R}^{n_t}} \hat{h}_t(x^{t-1}, x_t, s^t),
\end{aligned}$$

where

$$(E^{\nu_t} \check{h})(x^t, s^t) := \int_{\Xi_{t+1}} \check{h}_t(x^t, s^t, s_{t+1}) \nu_{t+1}(s^t, ds_{t+1}).$$

Proof. By Theorem 4.36, (BE) has a solution $(h_t)_{t=0}^T$. Assume that the claim holds for t. By the induction hypothesis,

$$\tilde{h}_{t-1}(x^{t-1},\omega) = \check{h}_{t-1}(x^{t-1},\xi^t(\omega))$$

almost surely everywhere. By Theorem 4.36, $\tilde{h}_{t-1}$ is L-bounded, so Corollary 4.35 gives

$$\begin{aligned} h_{t-1}(x^{t-1},\omega) &= (E_t\tilde{h}_{t-1})(x^{t-1},\omega) \\ &= (E^{\nu_{t-1}}\check{h}_{t-1})(x^{t-1},\xi^{t-1}(\omega)) \\ &= \hat{h}_{t-1}(x^{t-1},\xi^{t-1}(\omega)), \end{aligned}$$

so the claim holds for $t-1$. Clearly, the claim holds for $t = T$, so the proof follows by induction. □

4.3.2 Mathematical programming

Example 4.42 (Mathematical programming, continued)**.** *The problem in Example 3.45 admits a solution if there is a $p \in \mathcal{N}^\perp$, $\epsilon > 0$ and an $m \in L^1$ such that*

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

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

$$h_j(x) \leq 0 \quad j = 1,\dots,m \quad a.s.$$

and if

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

is a linear space.

Example 4.43 (Linear programming, continued)**.** *The problem in Example 3.47 admits a solution if there exists $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that*

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

for $\lambda \in [1-\epsilon, 1+\epsilon]$, and

$$\{x \in \mathcal{N} \mid c \cdot x \leq 0,\ Ax \leq 0\ a.s.\}$$

is a linear space.

4.3.3 Stochastic problems of Bolza

Example 4.44 (Stochastic problems of Bolza, dynamic programming)**.** *Consider again the problem*

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

from Example 3.49 and assume that

1. there exist $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that for every $\lambda \in (1-\epsilon, 1+\epsilon)$ there exist $y \in \mathcal{N}^1$ such that

$$EK_t^*(\lambda p_t + \Delta y_{t+1}, y_t) < \infty$$

for all t.

2. $\{x \in \mathcal{N} \mid \sum_{t=0}^T K_t^\infty(x_t, \Delta x_t) \le 0 \text{ a.s.}\}$ is a linear space.

Then the functions $V_t : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ defined by

$$\begin{aligned} \tilde{V}_T &:= 0, \\ V_t &:= E_t \tilde{V}_t, \\ \tilde{V}_{t-1}(x_{t-1}, \omega) &:= \inf_{x_t \in \mathbb{R}^d} \{K_t(x_t, \Delta x_t, \omega) + V_t(x_t, \omega)\}, \end{aligned} \tag{4.10}$$

are L-bounded convex normal integrands,

$$N_t(\omega) = \{x_t \in \mathbb{R}^d \mid K_t^\infty(x_t, x_t, \omega) + V_t^\infty(x_t, \omega) \le 0\}$$

are linear-valued,

$$\inf(\mathrm{SP}_u) = \inf_{x^t \in \mathcal{N}^t} E\left[\sum_{s=0}^t K_s(x_s, \Delta x_s) + V_t(x_t)\right] \quad t = 0, \dots, T,$$

there is an optimal solution $x \in \mathcal{N}$ with $x_t \perp N_t$ for every $t = 0, \dots, T$. An $x \in \mathcal{N}$ is optimal if and only if

$$x_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^d} \{K_t(x_t, \Delta x_t) + V_t(x_t)\} \quad a.s. \quad t = 0, \dots, T.$$

Proof. We apply Theorem 4.40 to

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

Condition 1 implies that there exists $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that for every $\lambda \in (1-\epsilon, 1+\epsilon)$, there exist $y \in \mathcal{N}^1$ and $m_t \in L^1$ with

$$K_t(x_t, \Delta x_t) \ge x_t \cdot (\lambda p_t + \Delta y_{t+1}) + \Delta x_t \cdot y_t - m_t.$$

Summing up the lower bounds,

$$h(x) \ge \lambda \sum_{t=0}^T x_t \cdot p_t - \sum_{t=0}^T m_t$$

so the assumptions of Theorem 4.40 are satisfied. It thus suffices to show that the solution of (BE) satisfies

$$
\begin{aligned}
h_t(x^t,\omega) &= \sum_{s=0}^{t} K_s(x_s,\Delta x_s,\omega) + V_t(x_t,\omega)\\
V_t(x_t) &\ge -x_t\cdot y_t - E_t\sum_{s=t}^{T} m_s
\end{aligned}
\tag{4.11}
$$

for every $t=0,\dots,T$. For $t=T$, (4.11) is obvious since $V_T=0$ by definition. Assuming that (**??**) holds for t, we get

$$
\begin{aligned}
\tilde h_{t-1}(x^{t-1},\omega) &= \inf_{x_t\in\mathbb{R}^d} h_t(x^{t-1},x_t,\omega)\\
&= \sum_{s=0}^{t-1} K_s(x_s,\Delta x_s,\omega) + \inf_{x_t\in\mathbb{R}^d}\{K_t(x_t,\Delta x_t,\omega)+V_t(x_t,\omega)\}\\
&= \sum_{s=0}^{t-1} K_s(x_s,\Delta x_s,\omega) + \tilde V_{t-1}(x_{t-1}).
\end{aligned}
$$

Combining Condition 1 with the lower bound in (4.11) gives

$$
\tilde V_{t-1}(x_{t-1}) \ge x_{t-1}\cdot(\lambda p_{t-1}-y_{t-1}) - m_{t-1} - E_t\sum_{s=t}^{T} m_s,
$$

so $\tilde V_{t-1}$ is L-bounded, while, by Theorem 3.24, the linearity of N_t implies that $\tilde V_{t-1}$ is a well-defined convex normal integrand. The lower bound of $\tilde V_{t-1}$ gives the lower bound for V_{t-1} in (4.11). Since for $s=0,\dots,t-1$, K_s is $\mathcal{F}_{t-1}$-measurable and L-bounded, Theorem 4.26 gives

$$
h_{t-1}(x^{t-1},\omega) = \sum_{s=0}^{t-1} K_s(x_s,\Delta x_s,\omega) + V_{t-1}(x_{t-1},\omega),
$$

so the claim follows by induction on t. □

The linearity condition in Example 4.44 holds, in particular, if $K_t^\infty \ge 0$ for all $t=0,\dots,T$ and the functions $K_t(0,\cdot)$ are almost surely inf-compact for all $t=0,\dots,T$. The lower bound condition will be discussed in Section **??**.

Example 4.45 (Stochastic problems of Bolza, independence)**.** *In the setting of Example 4.44, assume that each K_t is a $\mathcal{H}_t$-measurable normal integrand for a $\mathcal{H}_t\subseteq\mathcal{F}_t$ such that K_{t+1} is $\mathcal{H}_t$-conditionally independent of $\mathcal{F}_t$ in the sense of Section 4.1.5. Then $\tilde V_t$ is $\mathcal{H}_{t+1}$-measurable and*

$$
V_t = E^{\mathcal{H}_t}\tilde V_t.
$$

In particular, if $(\mathcal{H}_t)_{t=0}^{T+1}$ are mutually independent, then V_t are deterministic, by Theorem 4.15.

Proof. By definition, $V_T = 0$ is $\mathcal{H}_T$-measurable. If V_{t+1} is $\mathcal{H}_{t+1}$-measurable, then, by Theorem 3.24, $\tilde{V}_t$ is so as well, and $V_t = E^{\mathcal{H}_t}\tilde{V}_t$, by Theorem 4.15. Thus, the claim follows by induction. □

Example 4.46 (Stochastic problems of Bolza, Markovianity). *Let $(\xi_t)_{t=0}^T$ be a stochastic process such that each ξ_t takes values in a measurable space $(\Xi_t, \mathcal{A}_t)$. Let $\xi^t := (\xi_0, \dots, \xi_t)$ and $\Xi^t := \prod_{t=0}^t \Xi_t$ be equipped with $\mathcal{A}^t := \mathcal{A}_0 \otimes \dots \otimes \mathcal{A}_t$. In the setting of Example 4.44, assume that $\mathcal{F}_t = \sigma(\xi^t)$ and*

$$K_t(x_t, \Delta x_t, \omega) = \hat{K}_t(x_t, \Delta x_t, \xi^t(\omega))$$

for an Ξ-valued $(\mathcal{F}_t)_{t=0}^T$-process ξ and convex normal integrands $\hat{K}_t$ on $\mathbb{R}^d \times \mathbb{R}^d \times \Xi^t$. If ν_t is a regular ξ^t-conditional distribution of ξ_{t+1}, then the solution $(V_t)_{t=0}^T$ of (4.10) is given by

$$V_t(x_t, \omega) = \hat{V}_t(x_t, \xi^t(\omega)),$$

where $\hat{V}_t : \mathbb{R}^d \times \Xi^t \to \overline{\mathbb{R}}$ are convex normal integrands defined recursively by

$$\begin{aligned} \check{V}_T &= 0, \\ \hat{V}_t &= E^{\nu_t}\check{V}_t, \\ \check{V}_{t-1}(x_{t-1}, s^t) &= \mathrm{cl}_{x_{t-1}} \inf_{x_t}\{\hat{K}_t(x_t, \Delta x_t, s^t) + \hat{V}_t(x_t, s^t)\}. \end{aligned}$$

In particular, if

1. *$(\xi_t)_{t=0}^T$ is Markov in the sense that $\nu_t(s^t, \cdot)$ is independent of s^{t-1},*
2. *$\hat{K}_t$ is independent of s^{t-2},*

then $\hat{V}_t$ and $\check{V}_t$ are independent of s^{t-1}.

Proof. The proof follows by induction just like in the proof of Corollary **??**. □

Example 4.47 (Optimal stopping, continued). *Recall the problem*

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

Let S be the Snell envelope *of Z, i.e.,*

$$\begin{aligned} S_{T+1} &:= 0 \\ S_t &:= \max\{Z_t, E_t S_{t+1}\}. \end{aligned}$$

Then the optimum value of the optimal stopping problem coincides for all $t = 0, \dots, T$ with that of

$$\underset{x^t \in \mathcal{N}^t}{\text{maximize}} \ E\left[\sum_{s=0}^t Z_s \Delta x_s + E_t[S_{t+1}](1 - x_t)\right] \quad \text{subject to} \ \Delta x \geq 0, \quad x \leq 1 \ P\text{-a.s.}$$

In particular, the optimum value of the problem is ES_0. An $x \in \mathcal{N}$ is optimal if and only if

$$x_t \in \operatorname*{argmax}_{x_t \in \mathbb{R}^d}\{(Z_t - E_t[S_{t+1}])x_t \mid x_t \in [x_{t-1}, 1]\} \quad a.s. \quad t = 0, \dots, T.$$

In particular, there exists an optimal solution $x \in \mathcal{N}$ taking values in $\{0, 1\}$.

Proof. Exercise. Hint: This fits Bolza with

$$K_t(x_t, \Delta x_t) = \delta_{\mathbb{R}_+}(1 - x_t) - Z_t \Delta x_t + \delta_{\mathbb{R}_+}(\Delta x_{t-1})$$

□

In the above example, if Z is Markovian, then $E_t S_{t+1}$ is a function ψ_t of Z_t and the optimal strategy is to exercise when $Z_t \geq \psi_t(Z_t)$ (it is an exercise to fill in the details).

Example 4.48 (Stochastic problems of Bolza, homogeneity)**.** *Consider again the model of Example 4.44 and assume that there is a $\rho > 0$ such that*

$$K_t(\alpha x_t, \alpha \Delta x_t) = \alpha^\rho K_t(x_t, \Delta x_t), \quad t = 0, \dots, T$$

for all x and $\alpha > 0$. Then V_t has the same property for all t. Indeed, the claim clearly holds for $t = T$. Assuming it holds for t, we have

$$\begin{aligned}
\tilde{V}_{t-1}(\alpha x_{t-1}) &= \inf_{x_t}\{K_t(x_t, x_t - \alpha x_{t-1}) + V_t(x_t)\} \\
&= \inf_{x_t}\{K_t(\alpha x_t, \alpha \Delta x_t) + V_t(\alpha x_t)\} \\
&= \alpha^\rho \inf_{x_t}\{K_t(x_t, \Delta x_t) + V_t(x_t)\} \\
&= \alpha^\rho \tilde{V}_{t-1}(x_{t-1}).
\end{aligned}$$

By linearity of the conditional expectation, $V_{t-1}(\alpha x_{t-1}) = \alpha^\rho V_{t-1}(x_{t-1})$, so the claim follows by induction.

Since

$$\inf_{x_t}\{K_t(x_t, \Delta x_t) + V_t(x_t)\} = \inf_{\Delta x_t}\{K_t(\Delta x_t + x_{t-1}, \Delta x_t) + V_t(\Delta x_t + x_{t-1})\},$$

it is optimal to take $\Delta x_t = 0$, if

$$\partial V_t(x_{t-1}) + A^* \partial K_t(x_{t-1}, 0) \ni 0,$$

where $Ax := (x, x)$. The homogeneity property implies

$$\partial V_t(\alpha x_{t-1}) + A^* \partial K(\alpha x_{t-1}, 0) = \alpha^{\rho - 1}(\partial V_t(x_{t-1}) + A^* \partial K_t(x_{t-1}, 0))$$

so the region where $\Delta x_t = 0$ is optimal is a cone.

Similarly, if $K_t(\alpha x_t, \alpha \Delta x_t) = K_t(x_t, \Delta x_t) + \ln \alpha$ for all x and $\alpha > 0$, then $V_t(\alpha x) = V_t(x) + \ln \alpha$ for all t. This implies $\partial V_t(\alpha x) = \alpha \partial V_t(x)$ so the region where $\Delta x_t = 0$ is optimal is again a cone.

4.3.4 Stochastic control

Example 4.49 (Stochastic control, dynamic programming)**.** *Consider the problem*

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

and assume that

1. *There exists $p\in\mathcal{N}^\perp$ and $\epsilon>0$ such that for every $\lambda\in(1+\epsilon,1-\epsilon)$ there exist $y\in\mathcal{N}^1$ and $m\in L^1$ such that $A_t^*y_t, B_t^*y_t$, $y_t\cdot u_t$ integrable and*

$$L_t(X_t,U_t)\ge \lambda p_t\cdot(X_t,U_t)-\Delta(y_{t+1}\cdot X_{t+1})-m$$

for all feasible (X,U). Here $y_0:=y_{T+1}:=0$.

2. *$\{(X,U)\in\mathcal{N}\mid \sum_{t=0}^T L_t^\infty(X_t,U_t)\le 0,\ \Delta X_t = A_tX_{t-1}+B_tU_{t-1}\}$ is a linear space.*

Then $J_t:\mathbb{R}^N\times\Omega\to\overline{\mathbb{R}}$ and $I_t:\mathbb{R}^{N+M}\times\Omega\to\overline{\mathbb{R}}$ defined recursively by

$$\begin{aligned} I_{T+1} &:= 0\\ J_t(X_t) &:= \inf_{U_t\in\mathbb{R}^M}(L_t+E_tI_{t+1})(X_t,U_t),\\ I_t(X_{t-1},U_{t-1}) &:= J_t(X_{t-1}+A_tX_{t-1}+B_tU_{t-1}+u_t), \end{aligned}$$

are L-bounded convex normal integrands,

$$N_t(\omega):=\{U_t\in\mathbb{R}^M\mid L_t^\infty(0,U_t,\omega)+(E_tI_{t+1}^\infty)(0,U_t,\omega)\le 0\}$$

are linear-valued, there is an optimal solution $(X,U)\in\mathcal{N}$ with $U_t\perp N_t$ for every $t=0,\dots,T$, the optimum value of (SC) *coincides with that of*

$$\begin{array}{lll} \text{minimize} & E\left[\sum_{s=0}^{t} L_s(X_s,U_s)+J_t(X_t)\right] & \text{over } (X^t,U^t)\in\mathcal{N}^t, \\ \text{subject to} & \Delta X_s = A_sX_{s-1}+B_sU_{s-1}+u_s & s=1,\dots,t \end{array}$$

for all $t=0,\dots,T$, and an $U\in\mathcal{N}$ is optimal if and only if

$$U_t\in\operatorname*{argmin}_{U_t\in\mathbb{R}^M}\{L_t(X_t,U_t)+(E_tI_{t+1})(X_t,U_t)\}.$$

Condition 1 holds if and only if there exists $p\in\mathcal{N}^\perp$ and $\epsilon>0$ such that for every $\lambda\in(1+\epsilon,1-\epsilon)$ there exist $y\in\mathcal{N}^1$ and $m\in L^1$ such that $A_t^*y_t, B_t^*y_t$, $y_t\cdot u_t$ integrable and

$$EL_t^*(\lambda p_t-(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}))<\infty,$$

for all t. Here $y_0:=y_{T+1}:=0$, $A_{T+1}:=0$ and $B_{T+1}:=0$.

Proof. We apply Theorem 4.40 with $x_t = (X_t, U_t)$ and

$$h(x,\omega) = \begin{cases} \sum_{t=0}^T L_t(X_t, U_t, \omega) & \text{if } \Delta X_t = A_t(\omega)X_{t-1} + B_t(\omega)U_{t-1} + u_t(\omega) \quad t = 1,\dots,T, \\ +\infty & \text{otherwise.} \end{cases}$$

Summing up the lower bounds,

$$h(x) \geq \lambda \sum_{t=0}^T x_t \cdot p_t - (T+1)m$$

so the assumptions of Theorem 4.40 are satisfied. It thus suffices to show that the solution of (BE) satisfies

$$\begin{aligned} &h_t(x^t) = \begin{cases} \sum_{s=0}^t L_s(X_s, U_s) + (E_t I_{t+1})(X_t, U_t) & \text{if } \Delta X_s = A_s X_{s-1} + B_s U_{s-1} + u_s \quad s = 1,\dots,t, \\ +\infty & \text{otherwise,} \end{cases} \\ &J_t(X_t) \geq y_t \cdot X_t - (T-t)E_t m \end{aligned} \tag{4.12}$$

for every $t = 0,\dots,T$ and that J_t is a normal integrand. For $t = T$, the expression for h_T is clear. By Theorem 4.39, N_T is linear-valued, so, by Theorem 3.24, J_T is thus a normal integrand. The lower bound for J_T follows from Condition 1.

Assuming that (4.12) holds for t and that J_t is a normal integrand, we get,

$$\begin{aligned} \tilde{h}_{t-1}(x^{t-1}) &= \inf_{x_t \in \mathbb{R}^{n_t}} h_t(x^{t-1}, x_t) \\ &= \sum_{s=0}^{t-1} L_s(X_s, U_s) + J_t(X_{t-1} + A_t X_{t-1} + B_t U_{t-1} + u_t) \\ &= \sum_{s=0}^{t-1} L_s(X_s, U_s) + I_t(X_{t-1}, U_{t-1}) \end{aligned}$$

if $\Delta X_s = A_s X_{s-1} + B_s U_{s-1} + u_s$ for all $s = 1,\dots,t-1$, and $+\infty$ otherwise.

The lower bound in (4.12) gives

$$I_t(X_{t-1}, U_{t-1}) \geq y_t \cdot (X_{t-1} + A_t X_{t-1} + B_t U_{t-1} + u_t) - (T-t)E_t m$$

The integrability of $A_t^* y_t, B_t^* y_t$ and $y_t \cdot u_t$ imply that I_t is L-bounded, so Theorem 4.26 gives

$$h_{t-1}(x^{t-1}) = \sum_{s=0}^{t-1} L_s(X_s, U_s) + (E_{t-1} I_t)(X_{t-1}, U_{t-1}).$$

Combining the lower bound of I_t with Condition 1

$$\begin{aligned} &L_{t-1}(X_{t-1}, U_{t-1}) + E_{t-1} I_t(X_{t-1}, U_{t-1}) \\ &\geq \lambda p_{t-1} \cdot (X_{t-1}, U_{t-1}) - \Delta(y_t \cdot X_t) - m + E_{t-1}(y_t \cdot X_t - (T-t)E_t m), \end{aligned}$$

so, taking conditional expectations,

$$L_{t-1}(X_{t-1},U_{t-1})+E_{t-1}I_t(X_{t-1},U_{t-1})\geq y_{t-1}\cdot X_{t-1}-(T-t+1)E_{t-1}m.$$

Thus (4.12) holds for $t-1$. By Theorem 4.39, N_{t-1} is a linear space, so J_{t-1} is a normal integrand by Theorem 3.24. Thus the claim follows by induction on t.

By Fenchel's inequality, for feasible (X,U),

$$\begin{aligned}
&L_t(X_t,U_t)+L_t^*(\lambda p_t-(\Delta y_{t+1}+A_{t+1}^*y_{t+1},B_{t+1}^*y_{t+1}))\\
&\geq \lambda p_t\cdot(X_t,U_t)-(\Delta y_{t+1}+A_{t+1}^*y_{t+1})\cdot X_t-(B_{t+1}^*y_{t+1})\cdot U_t\\
&=\lambda p_t\cdot(X_t,U_t)+y_t\cdot X_t-y_{t+1}\cdot(X_t+A_{t+1}X_t+B_{t+1}U_t)\\
&=\lambda p_t\cdot(X_t,U_t)-\Delta(y_{t+1}\cdot X_{t+1})+y_{t+1}\cdot u_{t+1}
\end{aligned}$$

so Condition 1 is equivalent to the given dual formulation. □

The linearity condition in Example 4.49 is satisfied, in particular, if $L_t^\infty\geq 0$ and the functions L_0 and $L_t(0,\cdot)$ for $t=1,\dots,T$ are almost surely inf-compact. The lower bound condition will be analyzed when we apply the duality theory to the problem.

Example 4.50 (Stochastic control, independence)**.** *In the setting of Example 4.49, assume that A_t, B_t and L_t are $\mathcal{H}_t$-measurable for a $\mathcal{H}_t\subseteq\mathcal{F}_t$ such that A_{t+1}, B_{t+1} and L_{t+1} are $\mathcal{H}_t$-conditionally independent of $\mathcal{F}_t$ in the sense of Section 4.1.5. Then I_t is $\mathcal{H}_t$-measurable and*

$$E_tI_{t+1}=E^{\mathcal{H}_t}I_{t+1}.$$

In particular, if $(\mathcal{H}_t)_{t=0}^T$ are mutually independent, then E_tI_{t+1} are deterministic, by Theorem 4.15.

Proof. Assume that I_{t+1} is $\mathcal{H}_{t+1}$-measurable. By Theorem 4.15, $E_tI_{t+1}=E^{\mathcal{H}_t}I_{t+1}$ and thus, by Theorem 3.24, J_t is $\mathcal{H}_t$-measurable. By Theorem 3.254, this implies the $\mathcal{H}_t$-measurability of I_t. Clearly, I_T is $\mathcal{H}_T$-measurable, so the claim follows by induction. □

Example 4.51 (Stochastic control, Markovianity)**.** *Consider Example 4.49 and let*

$$\xi_t=(A_t,B_t,u_t).$$

Assume that $\mathcal{F}_t=\sigma(\xi^t)$ and let $\hat{L}_t$ be an $\mathcal{A}^t$-measurable normal integrand (see Corollary 3.17) such that

$$L_t(X_t,U_t,\omega)=\hat{L}_t(X_t,U_t,\xi^t(\omega)).$$

Then

$$J_t(X_t,\omega)=\hat{J}_t(X_t,\xi^t(\omega)),$$

where $\hat{J}_t : \mathbb{R}^N \times \Xi^t \to \overline{\mathbb{R}}$ *is defined recursively by*

$$\begin{aligned}\hat{I}_{T+1} &:= 0\\ \hat{J}_t(X_t, s^t) &:= \inf_{U_t \in \mathbb{R}^M} (\hat{L}_t + E^{\nu_t}\hat{I}_{t+1})(X_t, U_t, s^t),\\ \hat{I}_t(X_{t-1}, U_{t-1}, s^t) &:= \hat{J}_t(X_{t-1} + A_t X_{t-1} + B_t U_{t-1} + u_t, s^t).\end{aligned}$$

In particular, if

1. $(\xi_t)_{t=0}^T$ *is Markov in the sense that* $\nu_t(s^t, \cdot)$ *is independent of* s^{t-1}*,*
2. $\hat{L}_t$ *is independent of* s^{t-2}*,*

then $\hat{I}_t$ *and* $\hat{J}_t$ *are independent of* s^{t-1}*.*

4.3.5 Financial mathematics

The following improves on the existence criteria in Example 3.50.

Example 4.52 (Financial mathematics, continued)**.** *Consider again the optimal investment problem in Example 3.50 and assume that*

1. *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* $EV^*(\lambda^i y) < \infty$ *for two different* $\lambda^i \in \mathbb{R}$*,*
2. *the set*
$$\mathcal{L} = \{x \in \mathcal{N} \mid \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1} \geq 0,\ z_t \in D_t^\infty\ P\text{-a.s.}\}$$
is linear.

Then optimal solutions exist.

- *In the absence of portfolio constraints, the linearity of* $\mathcal{L}$ *is equivalent to the classical* no-arbitrage condition, *i.e., the terminal wealth* $\sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1}$ *is nonnegative almost surely if and only if it is zero almost surely. This is the case under the first condition if* Q *is equivalent to* P*.*

Proof. Exercise. □

Example 4.53 (Variance optimal hedging)**.** *Consider the problem*

$$\text{minimize}\, E\left(V_0 + \sum_{t=0}^{T-1} z_t \cdot \Delta s_{t+1} - u\right)^2$$

over $V_0 \in \mathbb{R}$ *and* $\mathcal{F}_t$*-measurable* $\mathbb{R}^d$*-valued functions* z_t*. Here* V_0 *is interpreted as an initial value of a self-financing trading strategy* z*. The problem has a solution*

Proof. Exercise. □

Example 4.54 (Financial mathematics, continued)**.** *Assuming that s is componentwise nonzero almost surely, we can write the optimal investment problem in Example 4.52 as a stochastic control problem*

$$\begin{array}{ll} \text{minimize} & EV(u - X_T) \quad \text{over} \quad U \in \mathcal{N}_{\tilde{D}}, \\ \text{subject to} & \Delta X_t = R_t \cdot U_{t-1}, \quad X_0 = 0. \end{array}$$

where $R_t^j := \Delta s_t^j / s_{t-1}^j$ is the rate of return on asset j, $U_t^j := s_t^j z_t^j$ is the amount of cash invested in asset j over the period $(t, t+1]$ and

$$\tilde{D}_t(\omega) = \{U \in \mathbb{R}^M \mid (U^j / s_t^j(\omega))_{j=1}^M \in D_t(\omega)\}.$$

We assume that V is not a constant function.

The problem fits Example 4.49 with $A_t = 0$, $B_t = R_t$, $u_t = 0$ and

$$\begin{aligned} L_0(X_0, U_0) &= \delta_{\{0\}}(X_0) + \delta_{\hat{D}_t}(U_0), \\ L_t(X_t, U_t) &= \delta_{\hat{D}_t}(U_t), \\ L_T(X_T) &= V(u - X_T). \end{aligned}$$

The inequalities in Condition 1 in Example 4.49 mean that

$$\begin{aligned} &\delta_{\hat{D}_{t-1}}(U_{t-1}) \geq \lambda((v_{t-1} - y_{t-1})X_{t-1} + y_t X_t + c_{t-1} \cdot U_{t-1}) - m_t, \\ &V(u - X_T) \geq -\lambda y_T X_T - m_{T+1} \end{aligned}$$

for feasible (X, U). Feasibility means that $X_t = X_{t-1} + R_t \cdot U_{t-1}$, so the inequalities become

$$\begin{aligned} (v_{t-1} - y_{t-1}) + y_t &= 0, \\ \delta_{\hat{D}_{t-1}}(U_{t-1}) &\geq \lambda(y_t R_t + c_{t-1}) \cdot U_{t-1} - m_t, \\ V(u - X_T) &\geq -\lambda y_T X_T - m_{T+1}. \end{aligned}$$

If there are no portfolio constraints, condition 1 is satisfied if and only if there is a martingale y such that $E[uy_T + V^*(\lambda y_T)] < \infty$ and $E_t[y_{t+1}R_{t+1}] = 0$ for all t, while the linearity condition 2 becomes the classical no arbitrage condition: If $X_T \geq 0$ almost surely, then $X_T = 0$ almost surely.

In the absense of portfolio constraints, the recursion in Example 4.49 becomes

$$\begin{aligned} J_T(X_T) &= V(u - X_T) \\ I_{t+1}(X_t, U_t) &= J_{t+1}(X_t + R_{t+1} \cdot U_t) \quad t = T-1, \dots, 0, \\ J_t(X_t) &= \inf_{U_t \in \mathbb{R}^M} (E_t I_{t+1})(X_t, U_t) \quad t = T-1, \dots, 0. \end{aligned} \tag{4.13}$$

Since V is nondecreasing noncostant, each J_t is nonincreasing.

Exercise 4.3.1. *In the setting of Example 4.54 without portfolio constraints, assume that* $u = 0$ *and* $V(u) = e^u$*. Prove (by induction), that*

$$J_t(X_t) = \alpha_t e^{-X_t},$$

where $\alpha_T = 1$ *and*

$$\alpha_t = \inf_{U_t} E_t[\alpha_{t+1} e^{-R_{t+1}\cdot U_t}].$$

Note that the amount of money invested into the risky assets does not depend on the level of wealth X*. Assume further that each* R_t *is independent of* $\mathcal{F}_{t-1}$*. Prove that* α_t *are constants and cost-to-go functions* J_t *are deterministic. Show, in the special case where* $R_t \sim N(\mu_t, \Sigma_t)$ *for deterministic* μ_t *and* Σ_t*,*

$$E[e^{-R_{t+1}\cdot U_t}] = e^{-U_t\cdot\mu_t + \frac{1}{2}U_t\cdot\Sigma_t U_t},$$

so optimal U_t *is* $\Sigma_t^{-1}\mu_t$*.*

Exercise 4.3.2. *In the setting of Example 4.54, making the change of variables* $p^j = U^j/X$ *(proportion of wealth invested in* j*), the last equation in* (4.13) *becomes*

$$J_t(X_t) = \inf_{p_t\in\mathbb{R}^M}(E_t I_{t+1})(X_t, X_t p_t).$$

Assuming that $u = 0$ *and* $V(u) = -\ln(-u)$*, prove (by induction), that*

$$J_t(X_t) = -\ln(X_t) + \alpha_t$$

where $\alpha_T = 0$ *and*

$$\alpha_t = \inf_{p\in\mathbb{R}^M} E_t[\alpha_{t+1} - \ln(1 + R_{t+1}\cdot p)].$$

Similarly, with the power function $V(u) = -(-u)^q$ *for* $q \in (0,1)$*, show that*

$$J_t(X_t) = -\alpha_t X^q,$$

where $\alpha_T = 1$ *and*

$$\alpha_t = -\inf_{p\in\mathbb{R}^M} E_t[-\alpha_{t+1}((1 + R_{t+1}\cdot p))^q].$$

For both the logarithmic and the power function, the proportions of wealth invested into the risky assets do not depend on the level of wealth X*. Again, if each* R_t *is independent of* $\mathcal{F}_{t-1}$*, then* α_t *are constants and the cost-to-go functions are deterministic.*

Example 4.55 (Financial mathematics, Markovianity)**.** *Let* $\xi_t = (s_t, R_t)$*... and* $u(\omega) = u(s_T(\omega))$*. The recursion* (4.13) *in Example 4.54 becomes*

$$\begin{aligned}
\hat{J}_T(X_T, \xi^T) &= V(u(s_T) - X_T)\\
\hat{I}_{t+1}(X_t, U_t, \xi^{t+1}) &= \hat{J}_{t+1}(X_t + R_{t+1}\cdot U_t, \xi^{t+1}) \quad t = T-1,\dots,0,\\
\hat{J}_t(X_t, \xi^t) &= \inf_{U_t\in\mathbb{R}^M}(E^{\nu_t}\hat{I}_{t+1})(X_t, U_t, \xi^t) \quad t = T-1,\dots,0.
\end{aligned} \tag{4.14}$$

If the returns R_{t+1} are independent of (S_t, R_t), we see, using the fact that $s_{t+1} = s_t(1 + R_{t+1})$, that $\hat{J}_t$ does not depend on R_t.

Many path-dependent claims can be handled similarly to European type claims $c(s_T)$ that depends only on the terminal values s_T of the underlyings s. For example, an Asian $c(\sum_t s_t)$ can be expressed using a new process $A_t = \sum_{t'=0}^t s_t$ with dynamics $A_{t+1} = A_t + (1+R_{t+1})s_t$, the running maximum $M_t = \sup_{t' \leq t} s_{t'}$ with $M_{t+1} = \max(M_t, (1 + R_{t+1})s_t)$. The joint process $\xi = (s, A, M, R)$ will be Markov as soon as R is so.

If in addition the claim $c(S_T) = 0$ and $R^0 = 0$, we see by induction that J_t are independent of ξ_t and

$$\begin{aligned} J_T(X_T) &= V(-X_T), \\ J_t(X_t) &= \inf_{U_t \in \mathbb{R}^d} \{E^{\xi_t} J_{t+1}(X_t + R_{t+1} \cdot U_t)\}. \end{aligned}$$

That is, the cost-to-go-functions J_t are deterministic functions of the state.

Example 4.56 (Kabanov's model)**.** *Consider Example 4.48 in the case where*

$$K_t(x_{t-1}, \Delta x_t) = \delta_{D_{t-1}}(x_{t-1}) + \delta_{C_t}(\Delta x_t).$$

If C and D are conical and that there is a $\rho > 0$ such that $k(\alpha x) = \alpha^\rho k(x)$ for all $x \in \operatorname{dom} k$ and $\alpha > 0$, Example 4.48 implies that $V_t(\alpha x, \omega) = \alpha^\rho V_t(x, \omega)$ and that it is optimal to take $\Delta x_t = 0$ (i.e. not to trade at time t), when x_{t-1} belongs to the "no-trade region"

$$S_t(\omega) := \{z \in \mathbb{R}^d \mid \partial V_t(z, \omega) + C_t^*(\omega) \ni 0\}$$

which is a $\mathcal{F}_t$-measurable closed convex cone-valued mapping. Similar conclusions hold when $k(\alpha x) = k(x) + \ln \alpha$ for all $x \in \operatorname{dom} k$ and $\alpha > 0$.

5 Duality

We will study probem (SP$_u$) within the conjugate duality framework; see Appendix. To this end, we will parametrize (**??**) by an $\mathbb{R}^m$-valued random variable. More precisely, we will study the *parametric stochastic optimization problem*

$$\text{minimize} \quad Ef(x, u) := \int_\Omega f(x(\omega), u(\omega), \omega) dP(\omega) \quad \text{over}\, x \in \mathcal{N}, \qquad (\text{SP}_u)$$

where f is a convex $\mathcal{F}$-normal integrand on $\mathbb{R}^n \times \mathbb{R}^m$ and the parameter u is an element of a space $\mathcal{U}$ of $\mathbb{R}^m$-valued random variables. In many applications, the parameter u is introduced only for the purposes dualization but in others, it has practical significance. For example, in problems of financial mathematics, the parameter can be used to describe the cash-flows of assets or liabilities. In stochastic control, u may be taken to be the driving noise in the system equation; see the applications below.

In order to apply the classical conjugate duality framework, we will first restrict x to the space L^∞ of essentially bounded adapted strategies endowed with a locally convex topology under which its dual space can be identified with L^1. Moreover, we will introduce another parameter by adding a nonadapted perturbation to x. Straightworward application of the functional analytic duality theory then yields a dual problem and optimality conditions for the restriction of (SP_u).

In order to recover the original optimization problem (SP_u) over the space $\mathcal{N}$ of all adapted strategies, we will show that its objective coincides with the lower semicontinuous hull of that of the restricted problem. The topological relaxation does not affect the optimum value, so the dual problem is unafected. Moreover, the validity of the pointwise optimality conditions remain valid by a simple application of Fenchel's inequality and the fundamental Lemma 3.39.

5.1 Integral functionals in duality

5.1.1 Dual spaces of random variables

We will assume that $\mathcal{U}$ and $\mathcal{Y}$ are decomposable spaces of random variables in separating duality?? under the bilinear form

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

Strictly speaking, the elements of the spaces are equivalence classes of random variables that coincide almost surely.

Lemma 5.1. *If $L^\infty \subset \mathcal{U}$ and $L^\infty \subset \mathcal{Y}$, then $\mathcal{U} \subseteq L^1$, $\mathcal{Y} \subseteq L^1$ and*

$$\sigma(L^1, L^\infty)|_{\mathcal{U}} \subseteq \sigma(\mathcal{U}, \mathcal{Y}), \quad \sigma(\mathcal{U}, \mathcal{Y})|_{L^\infty} \subseteq \sigma(L^\infty, L^1),$$
$$\tau(L^1, L^\infty)|_{\mathcal{U}} \subseteq \tau(\mathcal{U}, \mathcal{Y}), \quad \tau(\mathcal{U}, \mathcal{Y})|_{L^\infty} \subseteq \tau(L^\infty, L^1).$$

The L^0-topology on $\mathcal{U}$ is weaker than $\tau(\mathcal{U}, \mathcal{Y})$.

Proof. Let $u \in \mathcal{U}$ and $y^i = \mathrm{sign}(u^i)$ so that $\|u\|_{L^1} = E[u \cdot y]$, which is finite since $y \in L^\infty$. Thus $\mathcal{U} \subset L^1$, and, by symmetry, $\mathcal{Y} \subset L^1$. The inclusions $L^\infty \subseteq \mathcal{U} \subset L^1$ 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. Since $\tau(L^1, L^\infty)$ topology is the L^1-norm topology on L^1 and L^0-topology on L^1 is weaker than the norm topology, the last claim follows from $\tau(L^1, L^\infty)|_{\mathcal{U}} \subseteq \tau(\mathcal{U}, \mathcal{Y})$. □

We remark that the above lemma does not in fact use decomposability of $\mathcal{U}$ and $\mathcal{Y}$. We remark also that $\tau(L^1, L^\infty)$ is the L^1-norm topology. In particular, in the setting of Lemma 5.1, L^0-relative topology in $\mathcal{U}$ is weaker than $\tau(\mathcal{U}, \mathcal{Y})$.

Examples of decomposable spaces are L^p-spaces and their generalizations, Orlicz, Lorentz and Banach function Spaces, and locally convex function spaces. Orlicz spaces are already of interested in the context of optimal investment of expected utility. More general "rearrangement invariant Banach spaces" (Banach spaces with law invariant norms, i.e. $\|u\| = \|u'\|$ as soon as u and u' have the same distribution) are of interest, e.g., in the martingale theory and in optimization of law invariant risk measures.

A simple example of a nondecomposable space is the space of continuous functions (and many other spaces in analysis, like Sobolev spaces and spaces of bounded variation).

5.1.2 Conjugates of integral functionals

The following is a corollary of Theorem 3.29.

Theorem 5.2 (Jensen's inequality)**.** *Let h be a $\mathcal{G}$-measurable convex normal integrand such that Eh^* is proper on $E^{\mathcal{G}}\mathcal{Y}$ and that $E^{\mathcal{G}}\mathcal{U} \subset \mathcal{U}$ and $E^{\mathcal{G}}\mathcal{Y} \subset \mathcal{Y}$. Then*

$$Eh(E^{\mathcal{G}}u) \leq Eh(u)$$

for every $u \in \mathcal{U}$.

Proof. By the interchange rule in Theorem 3.29,

$$\begin{aligned} Eh(E^{\mathcal{G}}u) &= E[\sup_{y\in\mathbb{R}^m}\{(E^{\mathcal{G}}u)\cdot y - h^*(y)\}] \\ &= \sup_{y\in\mathcal{Y}\cap L^0(\mathcal{G})} E[(E^{\mathcal{G}}u)\cdot y - h^*(y)] \\ &= \sup_{y\in\mathcal{Y}\cap L^0(\mathcal{G})} E[u\cdot y - h^*(y)] \\ &\leq \sup_{y\in\mathcal{Y}} E[u\cdot y - h^*(y)] \\ &= Eh(u) \end{aligned}$$

for any $u \in \mathcal{U}$. □

Applying Theorem 5.2 to the indicator function of a random set gives the following.

Corollary 5.3. *If S is a $\mathcal{G}$-measurable closed convex random set, then $E^{\mathcal{G}}u \in L^1(S)$ for every $u \in L^1(S)$.*

The interchange rule (Theorem 3.29) gives directly formulas for the conjugates and subdifferentials of integral functionals on decomposable spaces.

Theorem 5.4. *Let h be a convex normal integrand. Then $Eh : \mathcal{U} \to \overline{\mathbb{R}}$ and $Eh^* : \mathcal{Y} \to \overline{\mathbb{R}}$ are conjugates of each other as soon as they are proper and then $y \in \partial Eh(u)$ if and only if $y \in \partial h(u)$ almost surely.*

Proof. Given $y \in \mathcal{Y}$, apply Theorem 3.29 to the normal integrand $h_y(u,\omega) := h(u,\omega) - u \cdot y(\omega)$. $\square$

Lemma 5.5. *Let h be a convex normal integrand such that $Eh : \mathcal{U} \to \overline{\mathbb{R}}$ is proper and lsc. Then*

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

Remark 5.6. *If Eh is lsc and proper on L^∞, then h is $\mathcal{N}_p^\perp$-bounded if and only if there exist $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that*

$$h^\infty(x,\omega) \geq x \cdot p(\omega) + \epsilon |x \cdot p(\omega)|.$$

Corollary 5.7. *Let $u \in \mathcal{U}$ and $y \in L^0$ satisfy $y \in \partial h(u)$ almost surely. If Eh is bounded from above on $u + \varepsilon B$ for some $\varepsilon > 0$ and $B \subset \mathcal{U}$ such that*

$$\left\{ y \in L^0 \mid \sup_{u \in B} E[u \cdot y] < \infty \right\} \subseteq \mathcal{Y},$$

then $y \in \mathcal{Y}$ and thus $y \in \partial Eh(u)$.

Proof. The pointwise subgradient condition implies

$$E[u' \cdot y] \leq Eh(u + u') - Eh(u) \quad \forall u' \in \mathcal{U},$$

so $\sup_{u' \in \varepsilon B} E[u' \cdot y] < \infty$. $\square$

5.2 Duality in stochastic optimization

In this section we will study the problem

$$\text{minimize} \quad Ef(x,u) := \int f(x(\omega), u(\omega), \omega) dP(\omega) \quad \text{over } x \in \mathcal{N}^\infty, \qquad (P_u^\infty)$$

where $\mathcal{N}^\infty := \mathcal{N} \cap L^\infty$. We embed the problem into the conjugate duality framework by introducing an additional parameter $z \in L^\infty := L^\infty(\Omega, \mathcal{F}, P, \mathbb{R}^n)$ and the extended real-valued convex function F on $L^\infty \times L^\infty \times \mathcal{U}$ defined by

$$F(x,z,u) = Ef(x,u) + \delta_{\mathcal{N}}(x - z).$$

We denote the associated optimum value function by

$$\varphi(z,u) := \inf_{x \in L^\infty} \{ Ef(x,u) \mid x - z \in \mathcal{N} \}.$$

The space L^∞ is in separating duality with $L^1 := L^1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ under the bilinear form

$$\langle z, p \rangle := E[z \cdot p].$$

The associated *Lagrangian* is the extended real-valued function L on $L^\infty \times L^1 \times \mathcal{Y}$ defined by

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

Recall that we denote the orthogonal complement of $\mathcal{N}^\infty$ by

$$\mathcal{N}^\perp := \{v \in L^1 \,|\, \langle x,v\rangle = 0 \ \forall x \in \mathcal{N}^\infty\}.$$

It is easily checked that $\mathcal{N}^\perp = \{v \in L^1 \,|\, E_t v_t = 0 \ \forall t = 0,\dots,T\}$. Since $\mathcal{N}^\infty$ is closed in probability, Lemma 5.1, implies that it is closed in $\tau(L^\infty, L^1)$ and thus, by convexity, also in $\sigma(L^\infty, L^1)$. Thus, by the bipolar theorem,

$$\mathcal{N}^\infty = \{x \in L^\infty \mid \langle x,p\rangle = 0 \quad \forall p \in \mathcal{N}^\perp\}.$$

The *Lagrangian integrand* $l : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ is defined by

$$l(x,y,\omega) := \inf_{u\in\mathbb{R}^m}\{f(x,u,\omega) - u\cdot y\}.$$

For any (x,y,ω), $l(\cdot,y,\omega)$ is convex and $l(x,\cdot,\omega)$ is upper semicontinuous and concave. For any $x \in L^0(\mathbb{R}^n)$ and $y \in L^0(\mathbb{R}^m)$, $\omega \mapsto l(x(\omega),y(\omega),\omega)$ is measurable by Theorems 3.6, 3.9 and 3.10.

Lemma 5.8. *If Ef is proper on $L^\infty \times \mathcal{U}$, then*

$$L(x,p,y) = \begin{cases} +\infty & \text{if } x \notin \operatorname{dom}_1 Ef, \\ E[l(x,y) - x\cdot p] & \text{if } x \in \operatorname{dom}_1 Ef \text{ and } p \in \mathcal{N}^\perp, \\ -\infty & \text{otherwise,} \end{cases}$$

$$F^*(v,p,y) = Ef^*(v+p,y) + \delta_{\mathcal{N}^\perp}(p),$$

$$\varphi^*(p,y) = Ef^*(p,y) + \delta_{\mathcal{N}^\perp}(p).$$

If Ef^* is proper on $L^1 \times \mathcal{Y}$, then F is closed.

Proof. By definition,

$$\begin{aligned} L(x,p,y) &= \inf_{(z,u)\in L^\infty\times\mathcal{U}} \{F(x,z,u) - \langle z,p\rangle - \langle u,y\rangle\} \\ &= \inf_{(z,u)\in L^\infty\times\mathcal{U}} \{E[f(x,u) - z\cdot p - u\cdot y] \,|\, x - z \in \mathcal{N}\} \\ &= \inf_{(z',u)\in L^\infty\times\mathcal{U}} \{E[f(x,u) - (x-z')\cdot p - u\cdot y] \,|\, z' \in \mathcal{N}\}, \end{aligned}$$

so the expression for L follows from interchange rule Theorem 3.29. When Ef is proper on $L^\infty \times \mathcal{U}$, the interchange rule gives

$$\begin{aligned} F^*(v,p,y) &= \sup_{x\in\mathcal{N}^\infty} \{\langle x,v\rangle - L(x,p,y)\} \\ &= \sup_{x\in L^\infty, z'\in L^\infty, u\in\mathcal{U}} \{\langle x,v\rangle - E[f(x,u) - (x-z')\cdot p - u\cdot y] \,|\, z' \in \mathcal{N}\} \\ &= \sup_{x\in L^\infty, z'\in L^\infty, u\in\mathcal{U}} \{E[x\cdot(v+p) + u\cdot y - f(x,u) - z'\cdot p] \,|\, z' \in \mathcal{N}\} \\ &= Ef^*(v+p,y) + \delta_{\mathcal{N}^\perp}(p), \end{aligned}$$

When Ef^* is proper on $L^1 \times \mathcal{Y}$, Ef is closed on $L^\infty \times \mathcal{U}$ by Theorem 5.4, so F is closed as a sum of closed functions. □

The *dual problem* of (P_u^∞) is that of minimizing φ^*; see Section 8.9. By Lemma 5.8, the dual problem can be written as

$$\text{minimize} \quad E[f^*(p,y) - u \cdot y] \quad \text{over} \quad (p,y) \in \mathcal{N}^\perp \times \mathcal{Y}. \tag{D_u}$$

Theorem 5.9. *If the primal and the dual are feasible, the following are equivalent*

(a) x solves (SP_u), (p,y) solves the (D_u) and there is no duality gap,

(b) x and (p,y) is a saddle-point of $(x,p,y) \mapsto L(x,y) + \langle u,y \rangle$,

(c) x is primal feasible, (p,y) is dual feasible and

$$(p,y) \in \partial f(x,u) \quad P\text{-a.s.}$$

(d) x is primal feasible, (p,y) is dual feasible and

$$p \in \partial_x l(x,y), \quad u \in \partial_y[-l](x,y) \quad P\text{-a.s.}$$

Proof. By ?? in 8.9, (a) and (b) are both equivalent to

$$(0,p,y) \in \partial F(x,0,u)$$

which means that $F(x,0,u) + F^*(0,p,y) = \langle u,y \rangle$. Using the expressions for F and F^* in Lemma 5.8 (primal-dual feasibility implies that F is proper), this means that $x \in \mathcal{N}^\infty$, $p \in \mathcal{N}^\perp$ and

$$Ef(x,u) + Ef^*(p,y) = E[x \cdot p] + E[u \cdot y].$$

Since $f(x,u,\omega) + f^*(p,y,\omega) \geq x \cdot p + u \cdot y$ for all $x,p \in \mathbb{R}^n$ and $u,y \in \mathbb{R}^m$, this is equivalent to (c). Equivalence of (c) and (d) follows from scenariowise application of ?? in 8.9. □

Remark 5.10. *In the deterministic setting, $\mathcal{N}^\perp = \{0\}$ so the dual problem becomes*

$$\text{minimize} \quad f^*(0,y) - u \cdot y \quad \text{over} \quad y \in \mathbb{R}^m \tag{D_u}$$

and the optimality conditions in Theorem 5.9 become

$$\begin{aligned} &0 \in \partial_x l(x,y), \\ &u \in \partial_y(-l)(x,y) \end{aligned}$$

which are the classical KKT conditions.

Remark 5.11 (Equivalent Lagrangians)**.** *As soon as Ef is proper on $\mathcal{N}^\infty \times L^\infty$, Fenchel's inequality implies that the negative part of $f^*(p,y)$ is integrable for all $(p,y) \in L^1 \times \mathcal{Y}$. Thus, the interchange rule Theorem 3.29 implies that for $p \in \mathcal{N}^\perp$ and $y \in \operatorname{dom}_2 Ef^*$,*

$$\begin{aligned}(\operatorname{cl}_x L)(x,p,y) &= \sup_{v \in L^1} \{\langle x, v\rangle - F^*(v,p,y)\} \\ &= \sup_{v \in L^1} E[x \cdot v - f^*(v+p,y)] \\ &= \sup_{v' \in L^1} E[x \cdot (v'-p) - f^*(v',y)] \\ &= E[\underline{l}(x,y) - x \cdot p],\end{aligned}$$

where

$$\underline{l}(x,y,\omega) = \sup_v \{\langle x, v\rangle - f^*(v,y,\omega)\}.$$

When $(p,y) \notin \mathcal{N}^\perp \times \operatorname{dom}_2 Ef^*$, we have $(\operatorname{cl}_x L)(x,p,y) = -\infty$.

When F is closed, L are $\operatorname{cl}_x L$ the upper and lower closures of the same equivalence class of saddle-functions; see ?? in 8.9. The functions

$$(x,p,y) \mapsto \begin{cases} E[l(x,y) - x \cdot p] & \text{if } p \in \mathcal{N}^\perp, \\ -\infty & \text{otherwise,} \end{cases}$$

and

$$(x,p,y) \mapsto \begin{cases} E[\underline{l}(x,y) - x \cdot p] & \text{if } p \in \mathcal{N}^\perp, \\ -\infty & \text{otherwise,} \end{cases}$$

lie between L and $\operatorname{cl}_x L$ so they are also in the same equivalence class.

Remark 5.12. *Define the saddle-function $(\operatorname{cl}_y El)$ on $L^\infty \times \mathcal{Y}$ by*

$$\begin{aligned}(\operatorname{cl}_y El)(x,y) &:= \inf_u \{Ef(x,u) - \langle u, y\rangle\} \\ &= \begin{cases} +\infty & \text{if } x \notin \operatorname{dom}_1 Ef, \\ E[l(x,y)] & \text{otherwise.} \end{cases}\end{aligned}$$

The notation $\operatorname{cl}_y El$ comes from the fact that, as soon as Ef is proper and closed, $\operatorname{cl}_y El$ is the upper closure of the saddle-function

$$(El)(x,y) := El(x,y).$$

This follows just like in Remark 5.11. Similarly, the lower closure of El is given by

$$\begin{aligned}(\operatorname{cl}_x El)(x,y) &:= \sup_v \{\langle x, v\rangle - Ef^*(v,y)\} \\ &= \begin{cases} E[\underline{l}(x,y)] & \text{if } y \in \operatorname{dom}_y Ef^*, \\ -\infty & \text{otherwise.} \end{cases}\end{aligned}$$

The equivalence class contains also the saddle-function $(x,y) \mapsto E\underline{l}(x,y)$.

5.3 Relaxation

While (P_u^∞) in Section 5.2 allows for a convenient dualization, there are interesting applications where the infimum in (P_u^∞) is not attained or there is a duality gap between (P_u^∞) and (D_u). In this section, we show that (SP_u) can be obtained from (P_u^∞) by taking the *lower semicontinuous (lsc) hull* of

$$F(x,z,u) := Ef(x,u) + \delta_{\mathcal{N}}(x-z) + \delta_{L^\infty}(x)$$

on $L^0 \times L^\infty \times \mathcal{U}$ where L^0 is endowed with the topology of *convergence in probability.* Recall that the *lsc hull* of a function f is the pointwise supremum of all lsc functions majorized by f. We denote the lsc hull of F by $\bar{F}$ and the associated value function by

$$\bar{\varphi}(z,u) = \inf_{x\in L^0} \bar{F}(x,z,u).$$

Clearly, $\varphi \geq \bar{\varphi}$, but the inequality can be strict; see Example 5.22 below. In the next section, we will give general conditions for closedness of $\bar{\varphi}$ and attainment of the infimum.

The following is a direct consequence of the definition of lsc hull.

Lemma 5.13. *We have $\varphi^* = \bar{\varphi}^*$. In particular, $\operatorname{cl}\varphi = \operatorname{cl}\bar{\varphi}$.*

Proof. Any function f is minorized by the constant function $\inf f$, which is continuous in any topology, so the infimum of a function is not affected by taking the lsc hull. It follows that

$$\begin{aligned}
\bar{\varphi}^*(p,y) &= \sup_{z,u}\{\langle z,p\rangle + \langle u,y\rangle - \bar{\varphi}(z,u)\} \\
&= \sup_{x,z,u}\{\langle z,p\rangle + \langle u,y\rangle - \bar{F}(x,z,u)\} \\
&= \sup_{x,z,u}\{\langle z,p\rangle + \langle u,y\rangle - F(x,z,u)\} \\
&= \varphi^*(p,y).
\end{aligned}$$

By the biconjugate theorem, $\operatorname{cl}\varphi = \operatorname{cl}\bar{\varphi}$. □

In particular, the dual problem associated with the relaxed problem coincides with that of (P_u^∞). The previous section gave explicit expressions for the latter. The following theorem gives optimality conditions for the relaxed problem. The *Lagrangian* $\bar{L} : L^0 \times L^1 \times \mathcal{Y} \to \overline{\mathbb{R}}$ associated with the relaxed problem is given by

$$\bar{L}(x,p,y) := \inf_{z\in L^\infty, u\in\mathcal{U}}\{\bar{F}(x,z,u) - \langle z,p\rangle - \langle u,y\rangle\}.$$

Theorem 5.14. *Assume that*

$$\bar{F}(x,z,u) = Ef(x,u) + \delta_{\mathcal{N}}(x-z)$$

and that the primal and the dual are feasible. The following are equivalent *(a)* x *solves* (SP_u)*,* (p,y) *solves the* (D_u) *and there is no duality gap,*

(b) x *and* (p,y) *is a saddle-point of*

$$(x,p,y)\mapsto \bar{L}(x,p,y)+\langle u,y\rangle,$$

(c) x *is primal feasible,* (p,y) *is dual feasible and*

$$(p,y)\in\partial f(x,u)\quad P\text{-}a.s. \tag{5.1}$$

(d) x *is primal feasible,* (p,y) *is dual feasible and*

$$p\in\partial_x l(x,y),\quad u\in\partial_y[-l](x,y)\quad P\text{-}a.s.$$

In particular, if $(p,y)\in\mathcal{N}^\perp\times\mathcal{Y}$ *is a feasible dual solution and there is no duality gap, then an* $x\in\mathcal{N}$ *solves* (SP_u) *if and only if* (5.2) *holds.*

Proof. The equivalence of (a) and (b) is simply a restatement of Remark **??** in 8.9. Let $x\in L^0$ and $(p,y)\in L^1\times\mathcal{Y}$ be feasible. By Fenchel's inequality,

$$f(x,u)+f^*(p,y)-u\cdot y\geq x\cdot p\quad P\text{-a.s.}$$

so

$$Ef(x,u)+E[f^*(p,y)-u\cdot y]\geq E[x\cdot p],$$

where $E[x\cdot p]=0$, by Lemma 3.39. The second inequality holds as an equality if and only if the former does which, in turn, is equivalent to the pointwise subdifferential conditions. □

The following lemma shows that the relaxed problem is precisely the original problem (SP_u).

Lemma 5.15. *Assume that* F *is proper. We have*

$$\bar{F}(x,z,u)=Ef(x,u)+\delta_{\mathcal{N}}(x-z)$$

as soon as the right side is lsc and proper on $L^0\times L^\infty\times\mathcal{U}$*.*

Proof. By assumption,

$$Ef(x,u)+\delta_{\mathcal{N}}(x-z)\leq\bar{F}(x,z,u)$$

for all $(x,z,u)\in L^0\times L^\infty\times\mathcal{U}$. To prove the converse, let $(\bar{x},\bar{z},\bar{u})\in L^0\times L^\infty\times\mathcal{U}$ be such that the left side is finite. Define $Q\ll P$ by

$$dQ/dP=\phi:=\frac{(1+|\bar{x}|^2)^{-1}}{E[(1+|\bar{x}|^2)^{-1}]}.$$

We have $\bar{x} \in \mathcal{X}$, where $\mathcal{X} := L^2(\Omega, \mathcal{F}, Q; \mathbb{R}^n)$ is paired with $\mathcal{V} = \mathcal{X}$ under the bilinear form

$$\langle x, v \rangle := E^Q[x \cdot v] = E[\phi x \cdot v].$$

Let $\tilde{F}$ be the lsc hull of F on $\mathcal{X} \times L^\infty \times \mathcal{U}$. Since the topology of $\mathcal{X} \times L^\infty \times \mathcal{U}$ is stronger than the relative topology of $L^0 \times L^\infty \times \mathcal{U}$,

$$\bar{F}(\bar{x}, \bar{z}, \bar{u}) \leq \tilde{F}(\bar{x}, \bar{z}, \bar{u}).$$

It thus suffices to show that

$$\tilde{F}(\bar{x}, \bar{z}, \bar{u}) \leq Ef(\bar{x}, \bar{u}).$$

By the biconjugate theorem, $\tilde{F} = F^{**}$. For any $(v, p, y) \in \mathcal{V} \times L^1 \times \mathcal{Y}$,

$$\begin{aligned}
F^*(v, p, y) &= \sup_{(x,z,u) \in L^\infty \times L^\infty \times \mathcal{U}} \{\langle x, v \rangle + \langle z, p \rangle + \langle u, y \rangle - Ef(x, u) \,|\, x - z \in \mathcal{N}\} \\
&= \sup_{(x,z,u) \in L^\infty \times L^\infty \times \mathcal{U}} \{E\left[x \cdot (\phi v) + z \cdot p + u \cdot y - f(x, u)\right] \,|\, x - z \in \mathcal{N}\} \\
&= \sup_{(x,z',u) \in L^\infty \times L^\infty \times \mathcal{U}} \{E\left[x \cdot (\phi v) + (x - z') \cdot p + u \cdot y - f(x, u)\right] \,|\, z' \in \mathcal{N}\} \\
&= Ef^*(\phi v + p, y) + \delta_{\mathcal{N}^\perp}(p),
\end{aligned}$$

where the last equality follows from the interchange rule Theorem 3.29. Hence,

$$\begin{aligned}
F^{**}(\bar{x}, \bar{z}, \bar{u}) &= \sup_{(v,p,y) \in \mathcal{V} \times \mathcal{N}^\perp \times \mathcal{Y}} \{\langle \bar{x}, v \rangle + \langle \bar{z}, p \rangle + \langle \bar{u}, y \rangle - Ef^*(\phi v + p, y)\} \\
&= \sup_{(v,p,y) \in \mathcal{V} \times \mathcal{N}^\perp \times \mathcal{Y}} E[\bar{x} \cdot (\phi v) + \bar{z} \cdot p + \bar{u} \cdot y - f^*(\phi v + p, y)] \\
&= \sup_{(v,p,y) \in \mathcal{V} \times \mathcal{N}^\perp \times \mathcal{Y}} E[\bar{x} \cdot (\phi v + p) + (\bar{z} - \bar{x}) \cdot p + \bar{u} \cdot y - f^*(\phi v + p, y)].
\end{aligned}$$

By Fenchel's inequality,

$$f(\bar{x}, \bar{u}) + f^*(\phi v + p, y) \geq \bar{x} \cdot (\phi v + p) + \bar{u} \cdot y.$$

Thus, when $Ef^*(\phi v + p, y) < \infty$, we get $[\bar{x} \cdot p]^+ \in L^1$ and, since $\bar{z} \cdot p \in L^1$, we have $[(\bar{x} - \bar{z}) \cdot p]^+ \in L^1$ and thus, by Lemma 3.39, $E[(\bar{z} - \bar{x}) \cdot p] = 0$. It follows that

$$F^{**}(\bar{x}, \bar{z}, \bar{u}) = \sup_{(v,p,y) \in \mathcal{V} \times \mathcal{N}^\perp \times \mathcal{Y}} E[\bar{x} \cdot (\phi v + p) + \bar{u} \cdot y - f^*(\phi v + p, y)] \leq Ef(\bar{x}, \bar{u}),$$

where the inequality follows from Fenchel's inequality. □

Example 5.16. *Instead of the relaxation done in this section, it might be natural to consider lsc hull of $(x, u) \mapsto Ef(x, u) + \delta_{\mathcal{N}^\infty}(x)$ on $\mathcal{N} \times \mathcal{U}$. However, it is possible that Ef is lsc on $\mathcal{N} \times \mathcal{U}$ and proper on $\mathcal{N}^\infty \times \mathcal{U}$, but the lsc hull does not coincide with Ef. Indeed, let $T = 1$, $n_0 = n_1 = 1$, trivial $\mathcal{F}_0$, and*

$$f(x, u, \omega) := \begin{cases} 0 & \text{if } x_1 = x_0 \xi(\omega), \\ +\infty & \text{otherwise.} \end{cases}$$

where $\xi \notin L^\infty$. Then the lsc-hull is simply $\delta_{\{0\}\times\mathcal{U}}$ which does not coincide with Ef.

Example 5.17. *This example shows that we can have*

$$\bar{F}(x,z,u) < Ef(x,u) + \delta_{\mathcal{N}}(x-z)$$

unless the right side is lsc. We continue Example 4.2.1: let $\alpha \in L^2(\mathcal{F}_0)$ and $v \in \mathcal{N}^\perp$ be such that $E\alpha v = \infty$ and

$$f(x,u,\omega) := \frac{1}{2}|x_0 - \alpha(\omega)|^2 + x_0 v(\omega).$$

Here

$$F(x,z,u) = E[\frac{1}{2}|x_0-\alpha|^2 + x_0 v] + \delta_{\mathcal{N}}(x-z) = E[\frac{1}{2}|x_0-\alpha|^2] + \langle z_0, v\rangle + \delta_{\mathcal{N}}(x-z)$$

so

$$\bar{F}(x,z,u) = E[\frac{1}{2}|x_0-\alpha|^2] + \langle z_0, v\rangle + \delta_{\mathcal{N}}(x-z).$$

Now $\bar{F}(\alpha,0,0) = 0$ while $Ef(\alpha,0) = \infty$.

Assumption 5.18. *There exists $(p,y) \in \operatorname{dom} Ef^* \cap (L^1 \times \mathcal{Y})$ such that*

$$Ef(x+z,u) = E[f(x+z,u) - x\cdot p]$$

for all $(x,z,u) \in \mathcal{N} \times L^\infty \times \mathcal{U}$.

Lemma 5.19. *Under Assumption 5.18, the function*

$$(x,z,u) \mapsto Ef(x,u) + \delta_{\mathcal{N}}(x-z)$$

is lsc on $L^0 \times L^\infty \times \mathcal{U}$.

Proof. Let (p,y) be as in Assumption 5.18 and

$$k(x,u,\omega) := f(x,u,\omega) - x\cdot p(\omega) - u\cdot y(\omega).$$

$$K(x,z,u) := Ek(x,u) + \delta_{\mathcal{N}}(x-z).$$

By Theorem 3.31, K is lsc on $L^0 \times L^\infty \times \mathcal{U}$. By Assumption 5.18,

$$Ef(x,u)+\delta_{\mathcal{N}}(x-z) = E[f(x,u)-(x-z)\cdot p+\delta_{\mathcal{N}}(x-z)] = K(x,u)+\langle u,y\rangle+\langle z,p\rangle,$$

so the left side is lsc, since it as a sum of lsc functions. □

Lemma 5.20. *A normal integrand f satisfies Assumption 5.18 if there exist $p \in \mathcal{N}^\perp$ and $\epsilon > 0$ such that*

$$\inf_{y\in\mathcal{Y}} Ef^*(\lambda p, y) < \infty$$

for $\lambda \in [1-\epsilon, 1+\epsilon]$.

Proof. The assumption gives $y, y' \in \mathcal{Y}$ such that $Ef^*(p,y) < \infty$ and $Ef^*((1+\epsilon)p, y') \leq \infty$. By Fenchel's inequality,

$$f(x,u,\omega) \geq x \cdot p(\omega) + u \cdot y(\omega) - f^*(p(\omega), y(\omega), \omega),$$
$$f(x,u,\omega) - x \cdot p(\omega) \geq \epsilon x \cdot p(\omega) + u \cdot y'(\omega) - f^*((1+\epsilon)p(\omega), y'(\omega), \omega).$$

Let $(x,z,u) \in \mathcal{N} \times L^\infty \times \mathcal{U}$. If either $Ef(x+z,u) < \infty$ or $E[f(x+z,u) - x \cdot p] < \infty$, the above inequalities and Lemma 3.39 give $E[x \cdot p] = 0$, so

$$Ef(x+z,u) = E[f(x+z,u) - x \cdot p],$$

which finishes the proof. $\square$

Remark 5.21. *The finiteness of the infimum in Lemma 5.20 means that there exist $y \in \mathcal{Y}$ and $\beta \in L^1$ such that*

$$f(x,u,\omega) \geq \lambda x \cdot p(\omega) + u \cdot y(\omega) - \beta(\omega) \quad \forall (x,u) \in \mathbb{R}^n \times \mathbb{R}^m$$

or, equivalently,

$$l(x, y(\omega), \omega) \geq \lambda x \cdot p(\omega) - \beta(\omega) \quad \forall x \in \mathbb{R}^n$$

This holds, in particular, if there exist $p \in \mathcal{N}^\perp$, $y \in \mathcal{Y}$ and $\epsilon > 0$ such that

$$Ef^*(\lambda p, y) < \infty$$

for $\lambda \in [1-\epsilon, 1+\epsilon]$. This means that

$$f(x,u,\omega) \geq x \cdot p + \epsilon |x \cdot p| + u \cdot y - \beta.$$

Example 5.22. *It is possible that $\bar{\varphi}$ is proper lsc while φ is not and thus that their subdifferentials do not coincide. Indeed, let*

$$f(x,u,\omega) = \frac{1}{2}(x_0 - 1)^2 + \delta_{\{0\}}(x_0 \xi(\omega) - x_1),$$

$\mathcal{F}_0$ be trivial and $\xi \in L^0(\mathcal{F}_1)$ with $\xi \notin L^\infty$. We get

$$\begin{aligned}
\bar{\varphi}(z,0) &= \inf_{x \in L^0} E[\frac{1}{2}(x_0 - 1)^2 + \delta_{\{0\}}(x_0 \xi - x_1)] + \delta_{\mathcal{N}}(x - z) \\
&= \inf_{x \in \mathcal{N}} E[\frac{1}{2}(x_0 + z_0 - 1)^2 + \delta_{\{0\}}((x_0 + z_0)\xi - x_1 - z_1)] \\
&= \inf_{x_0 \in \mathbb{R}} E[\frac{1}{2}(x_0 + z_0 - 1)^2] \\
&= E[\frac{1}{2}(z_0 - Ez_0)^2],
\end{aligned}$$

where the infimum at the second last line is attained at $x_0 = 1 - Ez_0$. It is an exercise to verify that $\partial \bar{\varphi}(z,0) = ((z_0 - Ez_0, 0), 0)$. Moreover, $\bar{\varphi}(0,0) = 0$.

On the other hand

$$\begin{aligned}\varphi(z,0) &= \inf_{x\in L^\infty} E[\frac{1}{2}(x_0-1)^2 + \delta_{\{0\}}(x_0\xi - x_1)] + \delta_{\mathcal{N}}(x-z)\\ &= \inf_{x\in \mathcal{N}^\infty} E[\frac{1}{2}(x_0+z_0-1)^2 + \delta_{\{0\}}((x_0+z_0)\xi - x_1 - z_1)]\\ &= \inf_{x_0\in\mathbb{R}} E[\frac{1}{2}(x_0+z_0-1)^2 + \delta_{L^\infty}((x_0+z_0)\xi)],\end{aligned}$$

so $\varphi(0,0) = 1$*. In particular,* φ *is not lsc at the origin, since* $\varphi(0,0) > \bar{\varphi}(0,0) = \operatorname{cl}\varphi(0,0)$.

For convenience, we summarize the findings of this section in the following. The *Lagrangian* $\bar{L} : L^0 \times L^1 \times \mathcal{Y} \to \overline{\mathbb{R}}$ associated with the relaxed problem is given by

$$\bar{L}(x,p,y) := \inf_{z\in L^\infty, u\in\mathcal{U}} \{\bar{F}(x,z,u) - \langle z,p\rangle - \langle u,y\rangle\}.$$

Theorem 5.23. *If* (SP_u) *is feasible,* F *is proper and Assumption 5.18 holds with* $p \in \mathcal{N}^\perp$*, then the following are equivalent*

(a) x *solves* (SP_u)*,* (p,y) *solves the* (D_u) *and there is no duality gap,*

(b) x *and* (p,y) *is a saddle-point of*

$$(x,p,y) \mapsto \bar{L}(x,p,y) + \langle u,y\rangle,$$

(c) x *is primal feasible,* (p,y) *is dual feasible and*

$$(p,y) \in \partial f(x,u) \quad P\text{-a.s.} \tag{5.2}$$

(d) x *is primal feasible,* (p,y) *is dual feasible and*

$$p \in \partial_x l(x,y), \quad u \in \partial_y[-l](x,y) \quad P\text{-a.s.}$$

(e) x *solves* (SP_u) *and* $(p,y) \in \partial\bar{\varphi}(0,u)$.

In particular, if $(p,y) \in \mathcal{N}^\perp \times \mathcal{Y}$ *is a feasible dual solution and there is no duality gap, then an* $x \in \mathcal{N}$ *solves* (SP_u) *if and only if* (5.2) *holds.*

By part (c) in the above theorem, optimal solutions, if they exist, are a subset of pointwise minimizers of $f(x,u) - x\cdot p - u\cdot y$. In this case, in particular, if the pointwise minimizer is unique, then it has to be adapted.

6 Duality of optimum values

6.1 Lower semicontinuity in L^0

Theorem 6.1. *Assume that f is bounded from below and that*

$$\{x \in \mathcal{N} \mid f^\infty(x,0) \leq 0\}$$

is a linear space. Then

$$\varphi(u) := \inf_{x\in\mathcal{N}} Ef(x,u)$$

is lower semicontinuous on $L^0(\mathbb{R}^m)$,

$$\varphi^\infty(u) = \inf_{x\in\mathcal{N}} Ef^\infty(x,u)$$

and the infimums are attained for every $u \in L^0(\mathbb{R}^m)$.

Proof. We may assume that $f \geq 0$. Let $h_u(x,\omega) = f(x,u(\omega),\omega)$. We have $h_u^\infty(x,\omega) = f^\infty(x,0,\omega)$ for every u, so, by Theorem 4.40, the infimum in $\varphi(u) = \inf_{x\in\mathcal{N}} Ef(x,u)$ is attained for every $u \in L^0$ by an $x \in \mathcal{N}$ with $x_t(\omega) \perp N_t(\omega)$ almost surely.

Since L^0 is a metric space, it suffices to prove sequential lower semicontinuity, which means that for any $\gamma \in \mathbb{R}$ and for any sequence $(u^\nu)_{\nu=1}^\infty$ such that

$$\varphi(u^\nu) \leq \gamma$$

and $u^\nu \to u$ in L^0, we have $\varphi(u) \leq \gamma$. We will prove this by establishing the existence of an $x \in \mathcal{N}$ such that $Ef(x,u) \leq \gamma$.

As observed at the beginning of the proof, there is for every ν an $x^\nu \in \mathcal{N}$ such that $x_t^\nu \perp N_t$ and

$$Ef(x^\nu, u^\nu) \leq \gamma.$$

Moreover, the mappings N_t are independent of u^ν. Since $Ef(x^\nu,u^\nu) \leq \gamma$, Lemma 8.33 gives a sequence of convex combinations

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

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

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

we have by convexity that

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

and $Ef(\bar{x}^\nu, \bar{u}^\nu) \le \gamma$. Moreover, we still have $\bar{x}^\nu_t \in N_t^\perp$ almost surely and $\bar{u}^\nu \to u$ almost surely.

Passing to a subsequence if necessary, we may assume that $\bar{u}^\nu \to u$ almost surely, so that the measurable function $\rho(\omega) := \sup_\nu |\bar{u}^\nu(\omega)|$ is almost surely finite. Each $(\bar{x}^\nu, \bar{u}^\nu)$ then belongs to the set

$$\mathcal{C} = \{(x,u) \in \mathcal{N} \times L^0 \,|\, (x,u) \in C \text{ a.s.}\},$$

where

$$C(\omega) = \{(x,u) \,|\, x_t \in N_t^\perp(\omega),\ |u| \le \rho(\omega),\ f(x,u,\omega) \le \phi(\omega)\}.$$

By Lemma 3.37, the sequence $(\bar{x}^\nu, \bar{u}^\nu)_{\nu=1}^\infty$ is almost surely bounded if

$$\{(x,u) \in \mathcal{N} \times L^0 \,|\, (x,u) \in C^\infty \text{ a.s.}\} = \{(0,0)\}. \tag{6.1}$$

We have

$$C^\infty(\omega) = \{(x,0) \,|\, x_t \in N_t^\perp(\omega),\ f^\infty(x,0,\omega) \le 0\}.$$

If $x \in \mathcal{N}$ is such that $f^\infty(x(\omega),0,\omega) \le 0$ then, by the last part of Theorem 4.39, we have $x_0 \in N_0$. The condition $x_0 \in N_0^\perp$ then implies that $x_0 = 0$. Repeating the argument for $t = 1,\dots,T$ gives (6.1) so $(\bar{x}^\nu, \bar{u}^\nu)_{\nu=1}^\infty$ is almost surely bounded.

By Lemma 8.33, there is a sequence $(\hat{x}^\nu, \hat{u}^\nu)_{\nu=1}^\infty$ of convex combinations of $(\bar{x}^\nu, \bar{u}^\nu)_{\nu=1}^\infty$ that converges almost surely to a point $(x, \hat{u})$, where necessarily $\hat{u} = u$ since $\bar{u}^\nu \to u$ almost surely. By convexity, $Ef(\hat{x}^\nu, \hat{u}^\nu) \le \gamma$ while, by Fatou's lemma,

$$Ef(x,u) \le \liminf_{\nu\to\infty} Ef(\hat{x}^\nu, \hat{u}^\nu).$$

Thus, φ is lsc and the infimum in its definition is attained.

As to the recession function, let $\bar{u} \in \operatorname{dom}\varphi$ and $\bar{x} \in \mathcal{N}$ be such that $\varphi(\bar{u}) = Ef(\bar{x}, \bar{u})$. We have

$$\varphi^\infty(u) = \sup_{\alpha>0} \frac{\varphi(\bar{u}+\alpha u) - \varphi(\bar{u})}{\alpha} = \sup_{\alpha>0} \inf_{x\in\mathcal{N}} Ef_\alpha(x,u),$$

where

$$f_\alpha(x,u) = \frac{f(\bar{x}+\alpha x, \bar{u}+\alpha u) - f(\bar{x},\bar{u})}{\alpha}.$$

Since $f_\alpha \le f^\infty$, we have $\varphi^\infty(u) \le \inf_{x\in\mathcal{N}} Ef^\infty(x,u)$.

To prove the converse, let $a > \sup_{\alpha>0} \inf_{x\in\mathcal{N}} Ef_\alpha(x,u)$. For every positive integer α, there is an $x^\alpha \in \mathcal{N}$ with $Ef_\alpha(x^\alpha,u) < a$ and $x^\alpha_t \in N_t^\perp$ for every t. The functions f_α are non-decreasing in α, so $Ef_1(x^\alpha,u) < a$ and we may proceed as in the first part of the proof to obtain a sequence of convex combinations $\tilde{x}^\alpha = \operatorname{co}\{x^{\alpha'} \mid \alpha' \ge \alpha\}$ and $\tilde{x}$ such that $\tilde{x}^\alpha \to \tilde{x}$ almost surely. By Fatou's lemma,

$$\begin{aligned} Ef^\infty(\tilde{x},u) &\le E[\liminf_\alpha f_\alpha(x^\alpha,u)] \\ &= \liminf_\alpha Ef_\alpha(x^\alpha,u) \\ &\le a, \end{aligned}$$

which completes the proof. ☐

6.2 Lower semicontinuity in $\mathcal{U}$

Theorem 6.2. *Assume that there exists* $(p, y) \in \operatorname{dom} Ef^* \cap (L^1 \times \mathcal{Y})$ *such that*

1. $Ef(x+z,u) = E[f(x+z,u) - x \cdot p]$ *for all* $(x, z, u) \in \mathcal{N} \times L^\infty \times \mathcal{U}$,
2. $\{x \in \mathcal{N} \mid f^\infty(x, 0) - x \cdot p \le 0 \ P\text{-a.s.}\}$ *is a linear space.*

Then

$$\varphi(z,u) = \inf_{x \in L^0}\{Ef(x,u) + \delta_{\mathcal{N}}(x-z)\}$$

is lower semicontinuous on $L^\infty \times \mathcal{U}$,

$$\varphi^\infty(z,u) = \inf_{x \in L^0} E\{f^\infty(x,u) + \delta_{\mathcal{N}}(z-z)\}$$

and the infimums are attained for every $(z, u) \in L^\infty \times \mathcal{U}$.

Proof. By change of variables,

$$\varphi(z,u) = \inf_{x \in \mathcal{N}} Ef(x+z,u).$$

The convex normal integrand

$$k(x,z,u,\omega) := f(x+z,u,\omega) + x \cdot p(\omega) + z \cdot p(\omega) + u \cdot y(\omega)$$

is bounded below by assumption. Applying Theorem 6.1,

$$\hat{\varphi}(z,u) := \inf_{x \in \mathcal{N}} Ek(x,z,u)$$

is lsc in $L^0 \times L^0$,

$$\hat{\varphi}^\infty(z,u) = \inf_{x \in \mathcal{N}} Ek^\infty(x,z,u)$$

and the infimums are attained. Since the topology on $L^\infty \times \mathcal{U}$ is stronger, $\hat{\varphi}$ is lsc on $L^\infty \times \mathcal{U}$. By the first assumption,

$$\varphi(z,u) := \hat{\varphi}(z,u) - \langle z, p\rangle - \langle u, y\rangle,$$

which proves all the claims. ☐

Corollary 6.3. *Assume that there exist* $p \in \mathcal{N}^\perp$ *and* $\epsilon > 0$ *such that*

1. $\inf_{y \in \mathcal{Y}} Ef^*(\lambda p, y) < \infty$ *for every* $\lambda \in [1-\epsilon, 1+\epsilon]$.
2. $\{x \in \mathcal{N} \mid f^\infty(x, 0) \le 0 \ P\text{-a.s.}\}$ *is a linear space.*

Then

$$\varphi(z,u) = \inf_{x\in L^0}\{Ef(x,u) + \delta_{\mathcal{N}}(x-z)\}$$

is lower semicontinuous on $L^\infty \times \mathcal{U}$*,*

$$\varphi^\infty(z,u) = \inf_{x\in L^0} E\{f^\infty(x,u) + \delta_{\mathcal{N}}(z-z)\}$$

and the infimums are attained for every $(z,u) \in L^\infty \times \mathcal{U}$*.*

Proof. We apply Theorem 6.2. By Lemma 5.20, Condition 1 implies that of Theorem 6.2. It suffices to show that

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

By Condition 1, there exists $y, y' \in \mathcal{Y}$ such that $Ef^*(p,y) < \infty$ and $Ef^*((1+\epsilon)p, y') < \infty$. By Fenchel's inequality,

$$\begin{aligned} f(x,u,\omega) &\ge x\cdot p(\omega) + u\cdot y(\omega) - f^*(p(\omega), y(\omega), \omega),\\ f(x,u,\omega) - x\cdot p(\omega) &\ge \epsilon x\cdot p(\omega) + u\cdot y'(\omega) - f^*((1+\epsilon)p(\omega), y'(\omega), \omega). \end{aligned}$$

Thus

$$\begin{aligned} f^\infty(x,0,\omega) &\ge x\cdot p(\omega),\\ f^\infty(x,0,\omega) - x\cdot p(\omega) &\ge \epsilon x\cdot p(\omega). \end{aligned}$$

If either $f^\infty(x,0) \le 0$ or $f^\infty(x,0) - x\cdot p \le 0$ almost surely, then $x\cdot p \le 0$ almost surely. Lemma 3.39 then implies $x\cdot p = 0$ almost surely. □

6.3 Applications

6.3.1 Financial mathematics

Example 6.4 (Financial mathematics, continued)**.** *Consider again the problem*

$$\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} x_t\cdot\Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}_D, \tag{6.2}$$

in Example 4.52 and the following dual problem

$$\text{minimize} \qquad E\left[V^*(y) + \sum_{t=0}^{T-1}\sigma_{D_t(\omega)}(p_t + y\Delta s_{t+1}) - uy\right] \quad \text{over } p \in \mathcal{N}^\perp, y \in \mathcal{Y}.$$

Under Condition 1 of Example 4.52, there is no duality gap and an $x \in \mathcal{N}$ *and* $(p,y) \in \mathcal{N}^\perp \times \mathcal{Y}$ *are primal and dual optimal, respectively, if and only they are feasible and*

$$\begin{aligned} u - \sum_{t_0}^{T-1} x_t\Delta s_{t+1} &\in \partial V^*(y),\\ p_t + y\Delta s_{t+1} &\in N_{D_t}(x_t) \quad t = 0,\dots,T. \end{aligned}$$

In particular, in the absence of portfolio constraints, y is a positive multiple of a martingale density.

Proof. The problem fits the duality framework with

$$f(x,u) = V(u - \sum_{t=0}^{T-1} x_t \Delta s_{t+1}) + \sum_{t=0}^{T-1} \delta_{D_t(\omega)}(x_t),$$

$$l(x,y) = -y \sum_{t=0}^{T-1} x_t \Delta s_{t+1} - V^*(y) + \sum_{t=0}^{T-1} \delta_{D_t(\omega)}(x_t),$$

$$f^*(v,y) = V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t(\omega)}(v_t + y\Delta s_{t+1}).$$

Thus Condition 1 of Example 4.52 implies that Assumption **??** is satisfied, so the claims follow from Theorem 5.23. □

Example 6.5 (Financial mathematics, reduced dual)**.** *In the setting of Example 6.4, assume that $s_t \in \mathcal{U}$ for all t and consider the reduced dual problem*

$$\text{minimize}\, E\left[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t}(E_t[y\Delta s_{t+1}]) - uy\right] \quad \text{over}\, y \in \mathcal{Y}.$$

If the reduced dual problem has a solution, then x is optimal if and only if it is feasible and there is y such that

$$u - \sum_{t_0}^{T-1} x_t \Delta s_{t+1} \in \partial V^*(y),$$
$$E_t[y\Delta s_{t+1}] \in N_{D_t}(x_t) \quad t = 0,\dots,T.$$

In particular, in the absence of portfolio constraints, y is a positive multiple of a martingale density.

Proof. Since the portfolio constraint D is adapted, Jensen's inequality in Theorem 5.2 gives, for every $(p,y) \in L^1 \times \mathcal{Y}$,

$$E[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t(\omega)}(p_t + y\Delta s_{t+1})] \geq E[V^*(y) + \sum_{t=0}^{T-1} \sigma_{D_t(\omega)}(E_t[p_t + y\Delta s_{t+1})]]$$

so for every $(p,y) \in L^1 \times \mathcal{Y}$,

$$\varphi^*(\pi^*(p,y)) \leq \varphi^*(p,y),$$

where $\pi^*(p,y) := (E_t[y\Delta s_{t+1}] - y\Delta s_{t+1}, y)$. □

6.3.2 Stochastic control

In this subsection we assume throughout the following.

Assumption 6.6. *We have*

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_T$$
$$\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T,$$

where $\mathcal{U}_t \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^N)$ and $\mathcal{Y}_t \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^N)$ are decomposable spaces in separating duality under the bilinear form $(x_t, y_t) \mapsto E[u_t \cdot y_t]$.

Example 6.7 (Stochastic control, continued)**.** *Consider again the problem*

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

together with the dual problem

$$\text{minimize} \quad E\left[\sum_{t=0}^{T} [L_t^*(p_t - (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1})) - y_t \cdot u_t]\right] \quad \text{over } p \in \mathcal{N}^\perp, y \in \mathcal{Y},$$

where $y_0 := y_{T+1} := 0$, $A_{T+1} := 0$ and $B_{T+1} := 0$.

Under assumption 1 of Example 4.49, an $x \in \mathcal{N}$ and $(p, y) \in \mathcal{N}^\perp \times \mathcal{Y}$ are primal and dual optimal, respectively, if and only if they are feasible and

$$\begin{aligned} p_t - (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) &\in \partial L_t(X_t, U_t), \\ \Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t \end{aligned} \tag{6.3}$$

almost surely.

Proof. This fits the general framework with $x = (X, U)$, $u = (u_1, \dots, u_T)$,

$$f(x, u) = \sum_{t=0}^{T} L_t(x_t) + \sum_{t=1}^{T} \delta_{\{0\}}(\pi \Delta x_t - C_t x_{t-1} - u_t),$$

$\pi = [I\ 0]$ and $C_t = [A_t\ B_t]$. Denoting $C_{T+1} = 0$, we get

$$\begin{aligned} l(x, y) &= \sum_{t=0}^{T} L_t(x_t) - \sum_{t=1}^{T} (\pi \Delta x_t - C_t x_{t-1}) \cdot y_t \\ &= \sum_{t=0}^{T} [L_t(x_t) + x_t \cdot (\pi^* \Delta y_{t+1} + C_{t+1}^* y_{t+1})] \\ &= \sum_{t=0}^{T} [L_t(X_t, U_t) + X_t \cdot (\Delta y_{t+1} + A_{t+1}^* y_{t+1}) + U_t \cdot B_t^* y_{t+1}] \end{aligned}$$

and

$$
\begin{aligned}
f^*(v,y) &= \sup_x \left\{ \sum_{t=0}^{T} \left(x_t \cdot (v_t - \pi^* \Delta y_{t+1} - C^*_{t+1} y_{t+1}) - L_t(x_t) \right) \right\} \\
&= \sum_{t=0}^{T} L^*_t(v_t - \pi^* \Delta y_{t+1} - C^*_{t+1} y_{t+1}) \\
&= \sum_{t=0}^{T} L^*_t(v_t - (\Delta y_{t+1} + A^*_{t+1} y_{t+1}, B^*_{t+1} y_{t+1}))
\end{aligned}
$$

□

Example 6.8 (Stochastic control, maximum principle)**.** *Let*

$$
\begin{aligned}
H_t(X_t, U_t, y_{t+1}) &:= L_t(X_t, U_t) + y_{t+1} \cdot (A_{t+1} X_t + B_{t+1} U_t), \\
\bar{H}_t(X_t, p_t, y_{t+1}) &:= \inf_{U_t \in \mathbb{R}^M} \{ H_t(X_t, U_t, y_{t+1}) - (X_t, U) \cdot p_t \}.
\end{aligned}
$$

The optimality conditions (6.3) *can be written as*

$$
\begin{aligned}
U_t &\in \operatorname*{argmin}_{U \in \mathbb{R}^M} \{ H_t(X_t, U, y_{t+1}) - (X_t, U) \cdot p_t \}, \\
-\Delta y_{t+1} &\in \partial_X \bar{H}_t(X_t, p_t, y_{t+1}), \\
\Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t
\end{aligned}
$$

If, for all $(X, U, y) \in \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^N$,

$$
\partial_{(X,U)} H(X, U, y) = \partial_X H(X, U, y) \times \partial_U H(X, U, y), \tag{6.4}
$$

the optimality conditions can be written as

$$
\begin{aligned}
U_t &\in \operatorname*{argmin}_{U \in \mathbb{R}^M} \{ H_t(X_t, U, y_{t+1}) - (X_t, U) \cdot p_t \}, \\
-\Delta y_{t+1} &\in \partial_X [H_t(X_t, U_t, y_{t+1}) - (X_t, U) \cdot p_t], \\
&\quad (= \partial_X H_t(X_t, U_t, y_{t+1}) - p_t) \\
\Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t
\end{aligned}
$$

almost surely. Condition (6.4) *holds, in particular, if* L *is of the form*

$$
L(X, U) = L^0(X, U) + L^1(X) + L^2(U),
$$

where L^0 *is differentiable.*

Proof. Exercise. Hint: use Lemma 8.29. □

Given an integrable process u

$$
u \mapsto {}^a u := (E_t u_t)_{t=0}^T.
$$

is the *adapted projection* of u.

Example 6.9 (Stochastic control, reduced dual). *In the setting of Example 6.7, assume that, for $t = 0, \dots, T$, columns of A_t and B_t are in $\mathcal{U}$ and EL_t is proper on L^∞. Given $y \in \mathcal{Y}$, Jensen's inequality in Theorem 5.2gives*

$$\inf_{p\in\mathcal{N}^\perp} E\sum_{t=0}^{T} L_t^*(p_t - (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}))$$
$$= E\sum_{t=0}^{T} L_t^*(-E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1})),$$

where the infimum is attained by

$$p_t = (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) - E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}). \quad (6.5)$$

Thus, $(p, y) \in \mathcal{N}^\perp \times \mathcal{Y}$ solves the dual problem if and only if y solves the reduced dual

$$\text{minimize} \quad E\left[\sum_{t=0}^{T} [L_t^*(-E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, E_t B_{t+1}^* y_{t+1})) - y_t \cdot u_t]\right] \quad \text{over } y \in \mathcal{Y}.$$

and p is given by (6.5). Clearly, the objective value of the reduced dual remains unchanged if we replace y by its adapted projection.

If the dual optimum is attained, then an x is optimal if and only if it is feasible and there is an adapted and dual feasible y such that

$$\begin{aligned} -E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) &\in \partial L_t(X_t, U_t), \\ \Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t \end{aligned} \quad (6.6)$$

almost surely.

Example 6.10 (Stochastic control, maximum principle with reduced dual). *Let*

$$\begin{aligned} H_t(X_t, U_t, y_{t+1}) &:= L_t(X_t, U_t) + E_t[y_{t+1} \cdot (A_{t+1} X_t + B_{t+1} U_t)], \\ \bar{H}_t(X_t, y_{t+1}) &:= \inf_{U_t \in \mathbb{R}^M} H_t(X_t, U_t, y_{t+1}). \end{aligned}$$

The conditions (6.6) can be written as

$$\begin{aligned} U_t &\in \operatorname*{argmin}_{U\in\mathbb{R}^M} H_t(X_t, U, y_{t+1}), \\ -E_t[\Delta y_{t+1}] &\in \partial_X \bar{H}_t(X_t, y_{t+1}), \\ \Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t \end{aligned}$$

almost surely. If, for all $(X, U, y) \in \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^N$,

$$\partial_{(X,U)} H(X, U, y) = \partial_X H(X, U, y) \times \partial_U H(X, U, y), \quad (6.7)$$

these can be written as

$$\begin{aligned}
U_t &\in \operatorname*{argmin}_{U\in\mathbb{R}^M} H_t(X_t, U, y_{t+1}),\\
-E_t[\Delta y_{t+1}] &\in \partial_X H_t(X_t, U_t, y_{t+1}),\\
\Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + u_t
\end{aligned}$$

almost surely. Condition (6.7) *holds, in particular, if L is of the form*

$$L(X, U) = L^0(X, U) + L^1(X) + L^2(U),$$

where L^0 is differentiable.

Proof. Exercise. □

Example 6.11 (Stochastic control, Cost-to-go function)**.** *Recall the dynamic programming recursion*

$$\begin{aligned}
I_{T+1} &:= 0\\
J_t(X_t) &:= \inf_{U_t\in\mathbb{R}^M} (L_t + E_t I_{t+1})(X_t, U_t),\\
I_t(X_{t-1}, U_{t-1}) &:= J_t(X_{t-1} + A_t X_{t-1} + B_t U_{t-1} + u_t)
\end{aligned}$$

from Example 4.49. If X, U and y satisfy the optimality conditions (6.6)*, then*

$$\begin{aligned}
(E_t y_t, 0) &\in \partial(L_t + E_t I_{t+1})(X_t, U_t)\\
E_t y_t &\in \partial J_t(X_t)
\end{aligned}$$

almost surely for all t.

Proof. By Lemma 8.29, the first subdifferential condition implies the second. If X, U and y satisfy the optimality conditions (6.3), then $p_T + (y_T, 0) \in \partial L_T(X_T, U_T)$, so $(E_T y_T, 0) \in \partial L_T(X_T, U_T)$. Assume now that $E_{t+1}y_{t+1} \in \partial J_{t+1}(X_{t+1})$. Combined with the optimality conditions (6.3), this gives, for every $(X', U') \in \mathbb{R}^N \times \mathbb{R}^M$,

$$L_t(X'_t, U'_t) \ge L_t(X_t, U_t) + [p_t - (\Delta y_{t+1} + A^*_{t+1} y_{t+1}, B^*_{t+1} y_{t+1})](X'_t - X_t, U'_t - U_t)$$

and

$$\begin{aligned}
&J_{t+1}(X'_t + A_{t+1} X'_t + B_{t+1} U'_t + u_t)\\
&\qquad \ge J_{t+1}(X_t + A_{t+1} X_t + B_{t+1} U_t + u_t)\\
&\qquad + E_{t+1}[y_{t+1}] \cdot (X'_t + A_{t+1} X'_t + B_{t+1} U'_t - X_t - A_{t+1} X_t - B_{t+1} U_t)
\end{aligned}$$

Taking conditional expectations and summing up gives

$$\begin{aligned}
&E_t[L_t(X'_t, U'_t) + I_{t+1}(X'_t, U'_t)]\\
&\qquad \ge E_t[L_t(X_t, U_t) + I_{t+1}(X_t, U_t)] + E_t[y_t] \cdot (X'_t - X_t).
\end{aligned}$$

or, equivalently,

$$
\begin{aligned}
L_t(X'_t, U'_t) &+ (E_t I_{t+1})(X'_t, U'_t) \\
&\geq L_t(X_t, U_t) + (E_t I_{t+1})(X_t, U_t) + E_t[y_t] \cdot (X'_t - X_t),
\end{aligned}
$$

which completes the proof.

□

7 Optimal investment and indifference pricing

Recall the optimal investment problem

$$
\text{minimize} \quad EV\left(u - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}\right) \quad \text{over} \quad x \in \mathcal{N}_0, \tag{ALM}
$$

where $\mathcal{N}_0 = \{x \in \mathcal{N} \mid x_T = 0 \ P\text{-a.s.}\}$ and V is a nondecreasing nonconstant convex function with $V(0) = 0$. For simplicity, we omit the portfolio constraints in this section.

The market satisfies the *no-arbitrage condition* if

$$
\left\{ \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \,\middle|\, \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} \geq 0, \ x \in \mathcal{N} \right\} = \{0\}. \tag{NA}
$$

We denote the nonnegative multiples of absolutely continuous martingale densities of s by $\mathcal{Q}$. It is an exercise to verify that a nonnegative $y \in \mathcal{Y}$ belongs to $\mathcal{Q}$ if and only if $E_t[y\Delta s_{t+1}] = 0$ for all $t = 0, \dots T-1$.

The following theorem is from the exercise section above.

Theorem 7.1. *Assume that the market satisfies* (NA) *and that there exists a martingale measure* $Q \ll P$ *of* s *such that* $EV^*(\lambda \frac{dQ}{dP}) < \infty$ *for two different* $\lambda \geq 0$*. If* $\frac{dQ}{dP} \in \mathcal{Y}$*, then the value function of* (ALM) *is closed in* $L^\infty \times \mathcal{U}$ *and* (ALM) *has a solution for all* $(0, u)$*.*

7.1 Pricing formulas for indifference prices

Given a current financial position $\bar{u}$, the *indifference price* of a claim u is given by

$$
\pi(u; \bar{u}) := \inf_{\alpha \in \mathbb{R}} \{\alpha \mid \bar{\varphi}(0, \bar{u} + u - \alpha) \leq \bar{\varphi}(0, \bar{u})\}.
$$

This is the least amount of cash the agent needs to cover the claim u without worsening her risk profile. The indifference price can be expressed

$$
\pi(u, \bar{u}) = \inf\{\alpha \mid c - \alpha \in \mathcal{A}(\bar{u})\},
$$

where the *acceptance set*

$$\mathcal{A}(\bar{u}) := \{u \in \mathcal{U} \mid \bar{\varphi}(0, u + \bar{u}) \leq \bar{\varphi}(0, \bar{u})\},$$

describes the claims that the agent finds "acceptable" given her risk preferences and her ability to trade in the market.

Theorem 7.2. *Assume that $\mathcal{F}_0$ is the trivial σ-field , φ is proper and lower semicontinuous, and that the price process is arbitrage-free. Then*

$$\pi(u; \bar{u}) = \sup_{y \in \mathcal{Y}} \{Euy - \sigma_{\mathcal{A}(\bar{u})}(y) \mid Ey = 1\}.$$

If $\inf \varphi < \varphi(\bar{u})$*, then*

$$\pi(u; \bar{u}) = \sup_{y \in \mathcal{Q}, \lambda > 0} \{E[uy - \lambda V^*(y/\lambda) - \bar{u}y] - \lambda \bar{\varphi}(\bar{u}) \mid Eq = 1\}.$$

Proof. We have

$$\begin{aligned}
\pi^*(y; \bar{u}) &= \sup_{u \in \mathcal{U}} \{E[uy] - \pi_s(u, \bar{u})\} \\
&= \sup_{\alpha \in \mathbb{R}, c \in \mathcal{U}} \{E[uy] - \alpha \mid c - \alpha \in \mathcal{A}(\bar{u})\} \\
&= \sup_{\alpha \in \mathbb{R}, \tilde{c} \in \mathcal{U}} \{E[(\tilde{u} + \alpha)y] - \alpha \mid \tilde{c} \in \mathcal{A}(\bar{u})\} \\
&= \sup_{\alpha \in \mathbb{R}, \tilde{c} \in \mathcal{U}} \{E[(\tilde{u}y + \alpha(y - 1)] \mid \tilde{u} \in \mathcal{A}(\bar{u})\} \\
&= \begin{cases} \sigma_{\mathcal{A}(\bar{u})}(y) & \text{if } Ey = 1, \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Thus the first formula follows from the biconjugate theorem since $\pi(\cdot, \bar{u})$ is proper and lower semicontinuous (an exercise).

The support function $\sigma_{\mathcal{A}(\bar{u})}$ satisfies

$$\begin{aligned}
\sigma_{\mathcal{A}(\bar{u})}(q) &= \sup_{u \in \mathcal{U}} \{\langle u, y \rangle \mid \varphi(u + \bar{u}) \leq \varphi(\bar{u})\} \\
&= \inf_{\lambda \geq 0} \sup_{u \in \mathcal{U}} \{\langle u, y \rangle - \lambda(\varphi(\bar{u} + u) - \varphi(\bar{u}))\} \qquad (7.1) \\
&= \inf_{\lambda \geq 0} \sup_{\tilde{u} \in \mathcal{U}} \{\langle \tilde{u} - \bar{u}, y \rangle - \lambda(\varphi(\tilde{u}) - \varphi(\bar{u}))\} \\
&= \inf_{\lambda \geq 0} \{\lambda \varphi^*(y/\lambda) - E[\bar{u}y] + \lambda \varphi(\bar{u})\}.
\end{aligned}$$

Here the second line is based on Lagrangian duality and the assumption that $\inf \varphi < \varphi(\bar{c})$ (exercise below) . Thus the result follows from Theorem 7.1 once we recall that taking the lsc-hull does not change the infimum. □

Exercise 7.1.1. *Prove that $\pi(\cdot, \bar{u})$ is proper and lsc. Hint: Prove sequential lower semicontinuous using the "direct method", get precompactness of "minimizing sequences" using recession analysis and recalling $\varphi^\infty(u) = \inf_{x \in \mathcal{N}} Ef^\infty(x, u)$.*

Exercise 7.1.2. *Prove the formula* (7.1)*. Hint: Apply conjugate duality to*

$$F(\alpha, u) = \langle u, y\rangle + \delta_{\mathbb{R}_-}(\varphi(u+\bar{u}) - \varphi(u) + \alpha)$$

where $\alpha \in \mathbb{R}$ *is paired with* $\lambda \in \mathbb{R}$*,* $\langle \alpha, \lambda\rangle := \alpha\lambda$*.*

Example 7.3 (Superhedging prices)**.** *For the superheding prices, we have* $V = \delta_{\mathbb{R}_-}$ *and* $\bar{c} = 0$*, so the above formula reduces to*

$$\pi(c) = \sup_{y \in \mathcal{Q}} \{E[uy] \mid Ey = 1\}.$$

We have recovered the pricing formula for superhedging prices in terms of absolutely continuous martingale measures of s*.*

8 Appendix

8.1 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 8.1. *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. □

8.2 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 8.2.1. *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 8.2.2. *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 8.2.3. *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 8.2.4. *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 8.2.5. *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* $2A' \subset A$.

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

Lemma 8.2. *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 8.3. *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 8.2.1 and 8.2.5. 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. □

8.3 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 8.3.1. *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 8.4. *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 8.5. *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 8.6. *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 8.4, 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 8.5. □

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

Proof. By Lemma 8.3, $\operatorname{cl} C$ 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

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

□

8.4 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 8.1 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 8.4.1. *A function g is convex if and only if the* strict epigraph

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

is convex.

Many algebraic operations also preserve convexity of functions.

Theorem 8.8. *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

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

   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

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

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

Exercise 8.4.2. *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*

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

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 8.9. *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 8.10. *If G is K-convex and g is convex such that $g(u_1) \le g(u_2)$ whenever $u_1 \in \operatorname{rge} G$ and $u_1 - u_2 \in K$, then $g \circ G$ is convex.*

Proof. Exercise. □

Exercise 8.4.3. *Let G a convex function on X and h a nondecreasing convex function on $\operatorname{rge} G$. Then $h \circ G$ is convex.*

Exercise 8.4.4. *The function $g(u) = \Pi_{i=1}^n u_i^{\lambda_i}$ is concave on $\mathbb{R}^n$, where $u = (u_1, \dots, u_n)$, $\lambda_i > 0$ and $\sum_{i=1}^n \lambda_i < 1$.*

Hint: When $\sum_{i=1}^n \lambda_i = 1$, apply Exercise 8.4.2. For the general case, combine with the composition rule.

8.5 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*

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

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 8.11. *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 8.12. *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 $x_1, x_2 \in \mathbb{R}^n$,

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

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 8.4.2. 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 8.4.2. 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 8.13. *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 8.14. *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 $u' + \lambda' u \in C$. □

Exercise 8.5.1. *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 8.15. *A convex set C in $\mathbb{R}^d$ is bounded if and only if $(\operatorname{cl} C)^\infty = \{0\}$.*

Proof. Exercise. □

Remark 8.16. *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 and $C^\infty = \{0\}$. 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 8.17. *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. □

Exercise 8.5.2. *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 8.18. *Assume that $g : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is a proper lsc convex function such that*

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

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 $\operatorname{lev}_{\le\beta} p \cap \mathbb{B}$ 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 $\operatorname{lev}_{\le\beta} f$ is bounded (and hence compact) for every β if and only if $\{x \mid f^\infty(x) \le 0\} = \{0\}$ (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 $\operatorname{lev}_{\le\beta} p \cap \mathbb{B} = \Pi(\operatorname{lev}_{\le\beta} \bar{g})$ where Π is the projection from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^m$. We have

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

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

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

8.6 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 8.19. *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 $g(u) \le M$ 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 8.20. *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 $u = \lambda u' + (1-\lambda)\bar{u}$. Now

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

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 8.21. *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 8.20 and lower semicontinuity of g, $\operatorname{int}\operatorname{lev}_\alpha g \neq \emptyset$. Thus continuity follows from Theorem 8.19. □

8.7 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

$$\{u \in U \mid \sup_{y \in K}\langle u, y\rangle < 1\} \tag{8.1}$$

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 8.6, 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 just the usual inner product.

Example 8.22. *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

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

Theorem 8.23 (Biconjugate theorem)**.** *Assume that g is a convex extended real-valued function on U such that $(\operatorname{lsc} g)(u) > -\infty$ for all u. Then $\operatorname{lsc} g = g^{**}$, i.e.,*

$$(\operatorname{lsc} g)(u) = sup_{y \in Y} \{\langle u, y \rangle - g^*(y)\}.$$

In particular, if g is a lsc proper convex function, then $g = g^{**}$, i.e., g has the dual representation

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

Proof. Recall that the closure of convex set in a locally convex space is the intersection of all closed half spaces containing the set. Applying this to the epigraph of $\operatorname{lsc} g$, we get that $\operatorname{lsc} g$ is the supremum of all affine continuous functions less than g, i.e.,

$$(\operatorname{lsc} g)(u) = \sup_{y \in Y, \alpha \in \mathbb{R}} \{\langle u, y \rangle - \alpha \mid \langle u, y \rangle - \alpha \le g(u)\}.$$

Let $y \in Y$ be fixed. Then $u \mapsto \langle u, y \rangle - \alpha$ is smaller than g if $\langle u, y \rangle - \alpha \le g(u)$ for all u, i.e.

$$\alpha \ge \sup_u \{\langle u, y \rangle - g(u)\} = g^*(y).$$

We got that, if $g^*(y) < \infty$, then the largest affine function less than g with the slope y is $u \to \langle u, y \rangle - g^*(y)$, whereas if $g^*(y) = \infty$, then there are no such affine functions. Since this is valid for any y, we get that $u \mapsto \sup\{\langle u, y \rangle - g^*(y)\}$ is the supremum of all affine functions less than g. Combining this with the first paragraph of the proof, we get the result. □

Exercise 8.7.1. *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

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

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 8.24. *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 8.25. *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 (8.1), $K^\circ \subset C$ 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 (8.1), 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 8.26. *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 $\alpha \in \mathbb{R}$,

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

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

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

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. □

8.7.1 Exercises

Exercise 8.7.2. *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 8.7.3. *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 8.7.4. *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.

8.8 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 8.8.1. *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 8.8.2. *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 8.8.3. *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 8.27. *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. □

Theorem 8.28. *For a proper convex g, we have $(g^*)^\infty = \sigma_{\mathrm{dom}\, g}$. If g is also lsc, then $g^\infty = \sigma_{\mathrm{dom}\, g^*}$.*

Proof. Exercise. □

8.9 Conjugate duality in optimization

Now we turn our attention to convex optimization problems. Given a vector space X and a locally convec topological vector space U, assume that F is a jointly convex extended real-valued function on $X \times U$ such that $F(x, \cdot)$ is lsc for all x. The *value function*

$$\varphi(u) := \inf_{x \in X} F(x, u)$$

gives the optimal value of the convex minimization problem

$$\text{minimize} \quad F(x, u) \quad \text{over}\, x \in X \tag{P_u}$$

as a function of u. The value function is always convex, so, when it is proper and lower semicontinuous, the biconjugate theorem gives

$$\inf_{x\in X} F(x,u) = \varphi(u) = \sup\{\langle u,y\rangle - \varphi^*(y)\}.$$

The optimization problem (P_u) is called the *primal problem* and

$$\text{maximize} \quad \langle u,y\rangle - \varphi^*(y) \quad \text{over}\, y\in Y \tag{D_u}$$

is called the *dual problem*. When φ is proper and lower semicontinuous, their optimal values coincide, and one says that there is *no duality gap*, or there is *strong duality* between (P_u) and (D_u). Even when φ is not lsc, the inequality $\varphi \ge \varphi^{**}$ means that, at least, the dual problem gives lower bounds to the optimal value of the primal problem. This is sometimes called as *weak duality* between (P_u) and (D_u).

Lemma 8.29. *Given a convex function F on $X\times U$, let*

$$\varphi(u) := \inf_x F(x,u).$$

We have $x \in \operatorname*{argmin}_x F(x,u)$ and $y\in\partial\varphi(u)$ if and only if $(0,y)\in\partial F(x,u)$.

Proof. The first two inclusions hold if and only if

$$F(x',u') \ge F(x,u) + \langle u'-u,y\rangle \quad \forall x',u'$$

which means that $(0,y)\in\partial F(x,u)$. □

A simple condition for the absence of duality gap and the existence of dual solutions is given by continuity of φ.

Theorem 8.30. *Assume that φ is proper and bounded from above in a neighborhood of u. Then the optimal values of (P_u) and (D_u) coincide, and (D_u) has a solution.*

Proof. Exercise. □

When also X is a LCTVS and F is jointly lsc, we have the following symmetry between the primal and the dual formulations.

Theorem 8.31. *Assume that $F : X\times\mathcal{U}\to\overline{\mathbb{R}}$ is proper lsc convex function on LCTVSs X and U and let V be the dual space of X. The conjugate of $x\mapsto F(x,u)$ is $(\operatorname{lsc}\gamma_u)$, where*

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

The function φ is lsc at u if and only if γ_u is lsc at the origin.

Proof. Exercise. □

Many fundamental theorems in duality theory correspond to minimax theorems of convex-concave functions. The function $L : X \times Y \to \overline{\mathbb{R}}$ defined by

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

is called the *Lagrangian* associated with F. It is a *convex-concave function* in the sense that $L(\cdot, y)$ is convex for all $y \in Y$ and $L(x, \cdot)$ is concave for all $x \in X$. We have

$$\inf_{x \in X} L(x, y) = -\varphi^*(y)$$

and

$$\sup_{y \in Y} \{L(x, y) + \langle u, y\rangle\} = F(x, u).$$

Denoting $L_u(x, y) := \langle u, y\rangle + L(x, y)$ we have the elementary inequalities

$$F(x, u) \geq L_u(x, y) \geq \langle u, y\rangle - \varphi^*(y)$$

and thus

$$\inf_x \sup_y L_u(x, y) = F(x, u) \geq \sup_y \{\langle u, y\rangle - \varphi^*(y)\} = \sup_y \inf_x L_u(x, y).$$

If we have an equality above, the common value is called a *saddle-value* of L_u. A pair $(x, y) \in X \times Y$ is called a *saddle-point* of L_u if

$$L_u(x, y') \leq L_u(x, y) \leq L_u(x', y) \quad \forall (x', y') \in X \times Y.$$

This means that the pair (x, y) satisfies the *Karush-Kuhn-Tucker* (KKT) condition

$$0 \in \partial_x L(x, y) \quad u \in \partial_y(-L)(x, y).$$

Theorem 8.32. *Assume that $F : X \times U \to \overline{\mathbb{R}}$ is jointly convex with $F(x, \cdot)$ lsc for every $x \in X$. The following are equivalent:*

1. $\inf(P_u) = \sup(D_u)$
2. *φ is lsc at u.*
3. *The saddle-value of L_u exists.*

Furthermore, the following are equivalent:

4 *x solves (P_u), y solves (D_u) and $\inf(P_u) = \sup(D_u)$.*

5 *The pair (x, y) satisfies the KKT-condition.*

Proof. Exercise. □

Exercise 8.9.1. *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.*

8.10 Facts from probability theory

Lemma 8.33 (Komlós' theorem)**.** *Let $(x^\nu)_{\nu=1}^\infty$ be a sequence in $L^0(\Omega,\mathcal{F},P;\mathbb{R}^n)$ which is either*

1. *bounded in L^1,*
2. *almost surely bounded in the sense that*
$$\sup_\nu |x^\nu(\omega)| < \infty \quad P\text{-a.s.}$$
3. *$\mathbb{R}_+$-valued and bounded in probability in the sense that*
$$\lim_{\alpha\to\infty}\sup_\nu P(x^\nu \geq \alpha) \to 0.$$

Then there is a sequence of convex combinations $\bar{x}^\nu \in \operatorname{co}\{x^\mu \,|\, \mu \geq \nu\}$ that converges almost surely to an $\mathbb{R}^n$-valued function ($\mathbb{R}_+$-valued in the last case).

Proof. Sufficiency of 2 and 3 imply that of 1. Indeed, if (x^ν) is bounded in L^1, then $(|x^\nu|)$ is bounded in probability, so there exists a subsequence of convex combinations for which $(|x^\nu|)$ converge almost surely. Then 2. implies the claim.

We prove next 2. Let $q(\omega) := \sup_\nu |x^\nu(\omega)|$ and Q be equivalent to P, defined by $dQ/dP := 1/(1+q^2)$. Then $E^Q|x^\nu|^2 \leq E[q^2/(1+q^2)] \leq 1$, so (x^ν) belongs to the unit ball of $L^2(Q)$. Since the unit ball is weakly compact, there exists, by convex compactness (Theorem **??**), a subsequence of convex combinations converging in $L^2(Q)$. Taking a further subsequence, we get Q-almost surely converging sequence. Since Q and P are equivalent, this sequence converges P-almost surely.

Let us now prove 3. The sequence

$$\gamma_\nu := \inf\{E[e^{-\bar{x}^\nu}] \mid \exists J : \bar{x}^\nu = \sum_{\nu'=\nu}^{J} \alpha^{\nu'} x^\nu, \sum \alpha^{\nu'} = 1, \alpha^{\nu'} \geq 0\}$$

increases to some $\gamma \leq 1$. Let $\bar{x}^\nu$ be a maximizing sequence so that $Ee^{-\bar{x}^\nu} \leq \gamma + 1/\nu$. For any $\epsilon > 0$, there is $\delta > 0$ with

$$e^{-(x+x')/2} \leq (e^{-x} + e^{-x'})/2 - \delta 1_{B_\epsilon}(x,x')$$

where $B_\epsilon := \{(x,x') \in \mathbb{R}^2_+ \mid |x-x'| \geq \epsilon, x \wedge x' \leq 1/\epsilon\}$. Since

$$\gamma_{\nu\wedge\nu'} \leq E[e^{-(\bar{x}^\nu - \bar{x}^{\nu'})/2}] \leq (E[e^{-\bar{x}^\nu}] + E[e^{-\bar{x}^{\nu'}}])/2 - \delta P((\bar{x}^\nu, \bar{x}^{\nu'}) \in B_\epsilon),$$

we have

$$\lim_{\nu,\nu'\to\infty} P((\bar{x}^\nu, \bar{x}^{\nu'}) \in B_\epsilon) = 0.$$

Since

$$E|e^{-\bar{x}^{\nu}} - e^{-\bar{x}^{\nu'}}| \leq \epsilon + 2e^{-1/\epsilon} + P((\bar{x}^{\nu}, \bar{x}^{\nu'}) \in B_{\epsilon}),$$

$(e^{-\bar{x}^{\nu}})$ is Cauchy in L^1. Passing to a subsequence, $(e^{-\bar{x}^{\nu}})$ converges almost surely. We leave it an exercise ?? to check that boundedness in probability implies that, outside a null set, $(e^{-\bar{x}^{\nu}})$ converges to a strictly positive random variable, which then implies the claim. $\square$