Convex stochastic optimization

Dynamic programming and duality in discrete time

Teemu Pennanen, King's College London
Ari-Pekka Perkkiö, Ludwig-Maximilian University of Munich

May 21, 2022

Introduction

- We study a general class of **convex stochastic optimization** (CSO) problems that covers
  - financial models with or without transaction costs, constraints or illiquidity effects.
  - convex stochastic control (including linear-quadratic),
  - classical SP models (LP, NLP,...) from operations research,
  - ...
- Combining convex analysis with stochastic analysis, we
  - simplify and extend the existing theory of CSO and of "functional analytic finance" to much more general models (constraints, transaction costs, illiquidity effects, semi-static trading strategies, ...).
  - resolve measurability problems in popular stochastic DP formulations.
  - derive duality and optimality conditions: stochastic KKT-conditions, costate variables and maximum principles, gradients of cost-to-go functions, calibration of martingale measures and prices systems to observed derivative prices, ...

Convex stochastic optimization

- Dantzig, Linear programming under uncertainty. Manag. Sci., 1955.
- Beale, On minimizing a convex function subject to linear inequalities. J. Roy. Statist. Soc. Ser. B., 1955.
- Bellman, Dynamic programming. Princeton U. Press, 1957.
- Rockafellar, Integrals which are convex functionals, Pacific J. Math., 1968.
- Rockafellar, Integral functionals, normal integrands and measurable selections, Lecture Notes in Math., Springer, 1976.
- Rockafellar and Wets, Nonanticipativity and L^1-martingales in stochastic optimization problems, Math. Progr., 1976
- Bertsekas and Shreve. Stochastic optimal control, The discrete time case, Academic Press Inc., 1978.
- Dalang, Morton, Willinger, Equivalent martingale measures and no-arbitrage in stochastic securities market models, Stoch. and Stoch. Rep., 1990.
- Bismut, *Conjugate convex functions in optimal stochastic control*, J. Math. Anal. Appl., 1973.

Convex stochastic optimization

Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^T, P)$, we study the optimization problem

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

where

- $\mathcal{N} = \{(x_t)_{t=0}^T \,|\, x_t \in L^0(\Omega, \mathcal{F}_t, P; \mathbb{R}^{n_t})\}$,
- $h : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a **convex normal integrand** $(n = n_0, \dots, n_T)$:
  - h is $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable,
  - $x \mapsto h(x, \omega)$ lsc and convex for P-almost every $\omega \in \Omega$,
- the integral is defined as $+\infty$ unless the positive part of the integrand is integrable.

Convex stochastic optimization

Example (Mathematical programming)

If

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

where f_j are convex normal integrands, affine for $j > l$, then h is a normal integrand and the problem becomes

$$\begin{array}{lll} \text{minimize} & Ef_0(x) & \text{over } x \in \mathcal{N}, \\ \text{subject to} & f_j(x) \le 0 & j = 1,\dots,l \ a.s., \\ & f_j(x) = 0 & j = l+1,\dots,m \ a.s. \end{array} \tag{MP}$$

The case $l = 0$ (no equality constraints) was studied by [Rockafellar and Wets, 1978] in the case of bounded strategies. When $f_j(\cdot,\omega)$ are all affine, we recover the stochastic LP studied since [Dantzig, 1955].

Convex stochastic optimization

Example (Optimal stopping)

If $n_t = 1$ for all t and

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

for an adapted real-valued process Z, then h is a normal integrand and the problem becomes

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

This is a convex relaxation of the **optimal stopping problem**. The relaxation does not affect the optimum value (works also in continuous time).

Convex stochastic optimization

Example (Stochastic control)

Let $x_t = (X_t, U_t)$ and

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

where L_t is $\mathcal{F}_t$-measurable convex normal integrand and A_t, B_t, W_t are $\mathcal{F}_t$-measurable random matrices/vectors. Then h is a normal integrand and the problem becomes

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

[Bertsekas and Shreve, 1979]: "First, in the usual stochastic programming model, the controls cannot influence the distribution of future states (see Olsen [01-03J], Rockafellar and Wets [R3-R4J], and the references contained therein). As a result, the model does not include as special cases many important problems such as, for example, the classical linear quadratic stochastic control problem."

Convex stochastic optimization

Example (Optimal investment)

Let $n_t = d$ for all t and

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

where s is an adapted **price process** and $V : \mathbb{R} \to \overline{\mathbb{R}}$ is convex. Then h is a normal integrand and the problem becomes

$$\underset{x \in \mathcal{N}}{\text{minimize}} \quad EV\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1}\right)$$

which is the problem of **optimal investment with liability** $c \in L^0$. This was studied by [Rásonyi and Stettner, 2005] for $c \in L^\infty$.

Convex stochastic optimization

Example (Semistatic hedging)

Consider the problem

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

where $\bar{J}$ is a finite set of quoted assets with payouts $\bar{c} = (\bar{c}_j)_{j \in \bar{J}}$. The function S_0 gives the cost of buying a portfolio in $\mathbb{R}^{\bar{J}}$ at the best available market prices. This fits the general format with the time index running from -1 to $T-1$, $\mathcal{F}_{-1} = \{\Omega, \emptyset\}$, $x_{-1} = \bar{x}$ and

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

Convex stochastic optimization

Example (Optimal investment in illiquid markets)

Let

$$h(x,\omega) = \begin{cases} \sum_{t=0}^{T} V_t(S_t(\Delta x_t,\omega) + c_t(\omega)) & \text{if } x_t \in D_t(\omega), \\ +\infty & \text{otherwise} \end{cases}$$

where

- $S_t : \mathbb{R}^d \times \Omega \to \mathbb{R}$ is such that $S_t(\cdot,\omega)$ are convex with $S_t(0,\omega) = 0$ and $S_t(x,\cdot)$ are $\mathcal{F}_t$-measurable,
- $\omega \mapsto D_t(\omega)$ is $\mathcal{F}_t$-measurable with $D_t(\omega)$ closed convex and $0 \in D_t(\omega)$,
- $V_t : \mathbb{R} \to \overline{\mathbb{R}}$ are convex.

Then h is a normal integrand and the problem becomes

$$\text{minimize} \quad E\sum_{t=0}^{T} V_t(S_t(\Delta x_t) + c_t) \quad \text{over } x \in \mathcal{N}_D.$$

This was studied in [Pennanen, 2014] and [Pennanen and Perkkiö, 2018].

Dynamic programming

- Bellman, Dynamic programming. Princeton U. Press, 1957.
- Rockafellar and Wets, *Nonanticipativity and L^1-martingales in stochastic optimization problems*, Math. Progr., 1976
- Evstigneev, Measurable selection and dynamic programming, Math. Oper. Res., 1976.
- Dynkin and Evstigneev, Regular conditional expectations of correspondences, Theory of Probability and its Applications, 1976.
- Bismut, Intégrales convexes et probabilités, J. Math. Anal. Appl., 1973.
- Bertsekas and Shreve. Stochastic optimal control, The discrete time case, Academic Press Inc., 1978.
- Carpentier, Chancelier, Cohen, De Lara, Stochastic multi-stage optimization, Springer, 2015.

Dynamic programming

- An extended real-valued random variable X is **quasi-integrable** if either X^+ or X^- is integrable.
- Given a quasi-integrable X and a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, there is an a.s. unique $\mathcal{G}$-measurable random variable $E^{\mathcal{G}} X$ (the $\mathcal{G}$-**conditional expectation of** X) such that
$$E\left[\alpha(E^{\mathcal{G}}X)\right] = E\left[\alpha X\right] \quad \forall \alpha \in L^\infty_+(\Omega, \mathcal{G}, P).$$

Definition

Given a normal integrand h, a $\mathcal{G}$-normal integrand $E^{\mathcal{G}}h$ is a $\mathcal{G}$-**conditional expectation of** h if
$$(E^{\mathcal{G}}h)(x) = E^{\mathcal{G}}[h(x)] \quad \text{a.s.}$$
for all $x \in L^0(\mathcal{G})$ such that $h(x)$ is quasi-integrable.

- If $h(x,\omega) = x \cdot v(\omega)$ for a $v \in L^1$, then $(E^{\mathcal{G}}h)(x,\omega) = x \cdot [E^{\mathcal{G}}v](\omega)$.
- The conditional expectation obeys natural **calculus rules**.
- If $E^{\mathcal{G}}X$ can be expressed in terms of a **probability kernel** then the same applies to $E^{\mathcal{G}}h$.

Dynamic programming

- Denote $x^t = (x_0, \dots, x_t)$, $n^t = n_0 + \cdots + n_t$, $E_t = E^{\mathcal{F}_t}$.
- 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 **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 $t = T, \dots, 0$.
- Provides dimension reduction (much more so with special structures), optimality conditions, computational techniques, existence of solutions, ...
- The above was analyzed by [Evstigneev, 1976] and [Rockafellar and Wets, 1976] in the case of uniformly compact feasible sets.

Dynamic programming

Theorem ("Verification theorem")

Assume that h is bounded from below, (SP) *is feasible and that the Bellman equations* (BE) *admit a solution $(h_t)_{t=0}^T$. Then an $\bar{x} \in \mathcal{N}$ solves* (SP) *if and only if*

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

Remark

In the setting of Theorem 8,

$$h_t(x^t) = \operatorname*{essinf}_{\tilde{x} \in \mathcal{N}} \left\{ E_t h(\tilde{x}) \mid \tilde{x}^t = x^t \right\} \quad \forall x^t \in L^0(\mathcal{F}_t)$$

Compare with [Rásonyi and Stettner, 2005].

Dynamic programming

Theorem (Existence of solutions)

Assume that h is convex, bounded from below and that

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

is a linear space. Then (BE) *has a unique solution* $(h_t)_{t=0}^T$*, where each* h_t *is a convex normal integrand.*

- Here h^∞ (**recession function** of h) is the normal integral given by

$$h^\infty(x,\omega) := \sup_{\lambda>0} \frac{h(\bar{x}+\lambda x) - h(\bar{x})}{\lambda}.$$

- Both theorems extend to integrands h not necessarily bounded from below. In the context of optimal investment, a sufficient condition is that the utility function has **reasonable asymptotic elasticity**.
- In the classical optimal investment model, the linearity condition is the **no-arbitrage** condition. In the presence of transaction costs, it becomes the **robust no-arbitrage** condition.

Dynamic programming

Example (Stochastic control)

Consider the problem

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

and assume that L_t are bounded from below and that

$$\{(X,U)\in\mathcal{N} \mid \sum_{t=0}^{T} L_t^{\infty}(X_t,U_t)\le 0,\ \Delta X_t = A_t X_{t-1} + B_t U_{t-1}\}$$

is a linear space (as e.g. in linear-quadratic control).

Dynamic programming

Example (Stochastic control, continued)

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

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

are convex normal integrands, optimal controls exists and are characterized by

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

Note that, if L_t are deterministic and A_t, B_t, W_t are independent of $\mathcal{F}_{t-1}$, then J_t are deterministic and optimal control is a function of the state only.

Dynamic programming

Example (Financial mathematics, continued)

Consider again the optimal investment problem and assume that

1. there exists a martingale measure $Q \ll P$ 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.

- If V has **reasonable asymptotic elasticity** then condition 1 holds if it holds merely for one λ (this means that the dual problem is feasible; see below).
- In the absence of portfolio constraints, linearity of $\mathcal{L}$ is equivalent to the classical **no-arbitrage condition**.

Duality

- Rockafellar, *Conjugate duality and optimization*, SIAM, 1974.
- Rockafellar and Wets, *Nonanticipativity and L^1-martingales in stochastic optimization problems*, Math. Progr., 1976
- Dalang, Morton, Willinger, *Equivalent martingale measures and no-arbitrage in stochastic securities market models*, Stoch. and Stoch. Rep., 1990.
- Schachermayer, *A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time*, Insurance Math. Econ., 1992.
- Pontryagin, Boltyanskii, Gamkrelidze, Mishchenko, *The mathematical theory of optimal processes*, Interscience Publishers John Wiley & Sons, 1962.
- Pennanen, *Convex duality in stochastic programming and mathematical finance*, Math. Oper. Res., 2011.
- Pennanen, Perkkiö, *Stochastic programs without duality gaps*, Math. Progr., 2012.
- Biagini, Pennanen, Perkkiö, *Duality and optimality in stochastic optimization and mathematical finance*, Journal of Convex Analysis, 2018.
- Bismut, *Conjugate convex functions in optimal stochastic control*, J. Math. Anal. Appl., 1973.

Conjugate duality

Conjugate duality studies parametric optimization problems of the form

$$\text{minimize} \quad F(x,u) \quad \text{over} \quad x \in X, \tag{P}$$

where the parameter u takes values in a locally convex vector space U which is in **separating duality** with another LCTVS Y. If F is **convex** on $X \times U$, then

- the **optimum value** $\varphi(u)$ is convex on U,
- the associated **Lagrangian**

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

  is convex-concave on $X \times Y$,
- the **conjugate** of φ can be expressed as

  $$\varphi^*(y) := \sup_{u \in U}\{\langle u,y\rangle - \varphi(u)\} = -\inf_{x \in X} L(x,y).$$

Conjugate duality

- If φ is **lower semicontinuous** (lsc), then $\varphi = \varphi^{**}$ so the optimum value equals that of the **dual problem**:

$$\text{maximize} \quad \langle u, y\rangle - \varphi^*(y)$$

- A $y \in Y$ solves the dual if and only if $y \in \partial\varphi(u)$, i.e.

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

- In this case an $x \in X$ solves (P) if and only if (x, y) is a **saddle point** of $(x, y) \mapsto L(x, y) - \langle u, y\rangle$.
- The saddle-point condition means that $(0, y) \in \partial F(x, u)$, or equivalently, that (x, y) satisfies the **KKT-conditions**

$$0 \in \partial_x L(x, y) \quad \text{and} \quad u \in \partial_y[-L](x, y).$$

- See [Rockafellar, 1974] for details and applications.

Conjugate duality

- The above covers all other convex optimization duality frameworks: LP duality, Lagrangian duality, Hamiltonian mechanics, convex optimal control, mass transportation and its generalizations, . . .
- Numerical algorithms are often based on the saddle-point formulation (interior point and gradient methods for constrained problems).
- The dual representation gives lower bounds for the optimum value much as in [Davis and Karatzas, "A deterministic approach to optimal stopping", 1994] or [Rogers, "Monte Carlo valuation of American options", 2002] for optimal stopping problems.

CSO duality

We will study the **parametric stochastic optimization problem**

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

where f is a convex $\mathcal{F}$-normal integrand on $\mathbb{R}^n \times \mathbb{R}^m$ and the parameters (z,u) vary in spaces $\mathcal{X}$ and $\mathcal{U}$ of $\mathbb{R}^n$- and $\mathbb{R}^m$-valued random variables.

- In many applications (and in [Rockefellar, 1974]), the parameter u is introduced only for the purposes dualization but, in others, it has practical significance.
- In problems of financial mathematics, u is typically the payout of a claim.
- In stochastic control, u is the additive noise.
- Note that the space $\mathcal{N} = \{(x_t)_{t=0}^T \,|\, x_t \in L^0(\Omega,\mathcal{F}_t,P;\mathbb{R}^{n_t})\}$ is not locally convex so we are slightly outside Rockafellar's framework. It turns out that the special structure of the problem allows us to get away with this.

CSO duality

In order to embed this to the conjugate duality framework, we assume that

- $\mathcal{X} \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ is a LCTVS in separating duality with another LCTVS $\mathcal{V} \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ under the bilinear form

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

- $\mathcal{U} \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ is a LCTVS in separating duality with another LCTVS $\mathcal{Y} \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ under the bilinear form

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

- all the spaces are **solid** and contain L^∞ (covers L^p, Orlicz, Lorentz, ...).

Let $\mathcal{X}_a^\perp := \{p \in \mathcal{V} \mid \langle x, p \rangle = 0 \ \forall x \in \mathcal{X} \cap \mathcal{N}\}$.

Theorem

If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$*, the dual problem can be written as*

$$\text{maximize} \quad \langle \bar{u}, y \rangle - Ef^*(p, y) \quad \text{over} \quad (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}. \qquad (D)$$

The variable $p \in \mathcal{X}_a^\perp$ describes the "price of information"; see [Rockafellar and Wets, 1976], [Back and Pliska, 1987], [Davis, 1992].

CSO duality

Theorem

If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (P) *and* (D) *are feasible, then following are equivalent*

1. x *solves* (P), (p,y) *solves* (D) *and* $\inf(\text{SP}) = \sup(D)$,
2. x *is feasible in* (SP), (p,y) *is feasible in* (D) *and*

$$(p,y) \in \partial f(x,\bar{u}) \quad P\text{-a.s.} \tag{2}$$

3. x *is feasible in* (SP), (p,y) *is feasible in* (D) *and*

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

where l *is the random saddle-function*

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

CSO duality

Clearly, the optimum value of the dual problem

$$\text{maximize} \quad \langle \bar{u}, y \rangle - Ef^*(p, y) \quad \text{over} \quad (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}. \qquad (D)$$

equals that of the **reduced dual**

$$\text{maximize} \quad \langle \bar{u}, y \rangle - g(y) \quad \text{over} \quad y \in \mathcal{Y}, \qquad (rD)$$

where

$$g(y) := \inf_{p \in \mathcal{X}_a^\perp} Ef^*(p, y).$$

- A pair (p, y) solves (D) iff y solves (rD) and p attains the infimum in $g(y)$.
- The infimum in the definition of g can be found analytically in many applications.

CSO duality

Consider again the problem

$$\begin{array}{lll} \text{minimize} & Ef_0(x) & \text{over } x \in \mathcal{N}, \\ \text{subject to} & H(x) \in K \end{array} \tag{MP}$$

If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$, then the dual problem can be written as

$$\begin{array}{ll} \text{maximize} & E \inf_{x \in \mathbb{R}^n} \{f_0(x) + y \cdot H(x) - x \cdot p\} \quad \text{over } (p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y} \\ \text{subject to} & \qquad\qquad y \in K^* \quad a.s. \end{array} \tag{D_{MP}}$$

Theorem

If $\operatorname{dom} Ef \cap (\mathcal{X} \times \mathcal{U}) \neq \emptyset$ *and* (MP) *and* (D_{MP}) *are feasible, then the following are equivalent*

1. x *solves* (MP), (p, y) *solves* (D_{MP}) *and* $\inf(MP) = \sup(D_{MP})$,
2. x *is feasible in* (MP), (p, y) *is feasible in* (D_{MP}) *and*

$$p \in \partial_x [f_0 + y \cdot H](x),$$
$$H(x) \in K, \quad y \in K^*, \quad y \cdot H(x) = 0$$

almost surely.

CSO duality

Theorem (Optimal stopping)

The optimum value of the optimal stopping problem

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

equals that of

$$\underset{y\in\mathcal{M}}{\text{minimize}} \quad y_0 \quad \text{subject to} \quad y \geq R,$$

where $\mathcal{M}$ is the set of **martingales**. *The optimal $\tau \in \mathcal{T}$ and $y \in \mathcal{M}$ are characterized by $y \geq R$ and $y_\tau = R_\tau$.*

CSO duality

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

Theorem (Optimal control)

If $\operatorname{dom} Ef \cap \mathcal{X} \times \mathcal{U} \neq \emptyset$, *the dual problem control problem can be written as*

$$\begin{array}{ll} \text{maximize} & E\left[\sum_{t=1}^{T} W_t \cdot y_t - \sum_{t=0}^{T} L_t^*(p_t - (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}))\right] \\ \text{over} & (p, y) \in \mathcal{X}_a^{\perp} \times \mathcal{Y}. \end{array} \tag{D_{OC}}$$

If, in addition, both (OC) *and* (D_{OC}) *are feasible, then the following are equivalent*

1. (X, U) *solves* (OC), (p, y) *solves* (D_{OC}) *and there is no duality gap,*
2. $(X, U) \in \mathcal{N}$, $(p, y) \in \mathcal{X}_a^{\perp} \times \mathcal{Y}$ *and, almost surely for all* t,

$$\begin{aligned} p_t - (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) &\in \partial L_t(X_t, U_t), \\ \Delta X_t &= A_t X_{t-1} + B_t U_{t-1} + W_t. \end{aligned}$$

CSO duality

The scenariowise optimality conditions in OC mean that (X, U) satisfies the system equations and that

$$\begin{aligned} U_t &\in \underset{U_t \in \mathbb{R}^M}{\operatorname{argmin}} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\}, \\ -\Delta y_{t+1} &\in \partial_X \bar{H}_t(X_t, p_t, y_{t+1}), \end{aligned}$$

where the **Hamiltonian** H_t is defined by

$$H_t(X_t, U_t, y_{t+1}) := L_t(X_t, U_t) + y_{t+1} \cdot (A_{t+1} X_t + B_{t+1} U_t).$$

and

$$\bar{H}_t(X_t, p_t, y_{t+1}) := \inf_{U_t \in \mathbb{R}^M} \{H_t(X_t, U_t, y_{t+1}) - (X_t, U_t) \cdot p_t\}.$$

- This is the **stochastic maximum principle** for general convex control problems.
- The optimal **costate** y thus solves a **backward stochastic difference inclusion**.
- BSDEs were originally introduced by [Bismut, 1973] who analyzed continuous time models in Rockafellar's duality framework.
- It can be shown that an optimal costate y is also a subgradient of the cost-to-go function in dynamic programming.

CSO duality

Remark (Reduced dual)

Assume that each L_t is $\mathcal{F}_t$-measurable and that each EL_t is proper on $\S \times \mathcal{C}$. Then, under mild conditions on the spaces, the optimum value of the dual problem (D_{OC}) equals that of the reduced dual problem

$$\underset{y \in \mathcal{Y}_a}{\text{maximize}} \qquad E\left[\sum_{t=1}^{T} W_t \cdot y_t - \sum_{t=0}^{T} [L_t^*(-E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, E_t B_{t+1}^* y_{t+1}))]\right].$$

A pair (p, y) solves (D_{OC}) if and only $(E_t y_t)_{t=0}^T$ solves the reduced dual and

$$p_t = (\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) - E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}).$$

If (D_{OC}) has a solution, then an x is optimal if and only if it is feasible and there is a y feasible in the reduced dual such that

$$-E_t(\Delta y_{t+1} + A_{t+1}^* y_{t+1}, B_{t+1}^* y_{t+1}) \in \partial L_t(X_t, U_t)$$

almost surely.

CSO duality

$$\text{minimize} \qquad EV\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \bar{c}\cdot\bar{x} + S_0(\bar{x})\right) \text{ over } x \in \mathcal{N},\ \bar{x} \in \mathbb{R}^J \qquad (SSH)$$

Theorem

If S_0 is positively homogeneous, the dual problem can be written as

$$\begin{aligned} &\text{maximize} && E\left[cy - V^*(y)\right] \quad \text{over } p \in \mathcal{X}_a^\perp, y \in \mathcal{Y} \\ &\text{subject to} && p_{-1} + y\bar{c} \in y \operatorname{dom} S_0^* \\ & && p_t + y\Delta s_{t+1} = 0 \quad t = 0, \dots, T. \end{aligned} \qquad (D_{SSH})$$

If both (SSH) and (RD_{SSH}) are feasible, then the following are equivalent

1. x *solves* (SSH), (p, y) *solves* (RD_{SSH}) *and there is no duality gap,*
2. $x \in \mathcal{N}$, $(p, y) \in \mathcal{X}_a^\perp \times \mathcal{Y}$ *and, almost surely*

$$\begin{aligned} p_{-1} + y\bar{c} &\in \partial(yS_0)(x), \\ p_t + y\Delta s_{t+1} &= 0 \quad t = 0, \dots, T, \\ y &\in \partial V\left(c - \sum_{t=0}^{T-1} x_t \cdot \Delta s_{t+1} - \bar{c}\cdot\bar{x} + S_0(\bar{x})\right). \end{aligned}$$

CSO duality

Minimizing over $p \in \mathcal{X}_a^\perp$ gives the **reduced dual**

$$\begin{array}{lll} \text{maximize} & E\left[cy - V^*(y)\right] \quad \text{over} \;\; y \in \mathcal{Y} & \\ \text{subject to} & E[y\bar{c}] \in E[y] \operatorname{dom} S_0^*, & (RD_{SSH}) \\ & E_t[y\Delta s_{t+1}] = 0 \quad t = 0, \dots, T. & \end{array}$$

If $Ey > 0$, the constraints can be written as

$$\begin{aligned} E^Q \bar{c} &\in \operatorname{dom} S_0^*, \\ E_t^Q[\Delta s_{t+1}] &= 0 \end{aligned}$$

where $dQ/dP := y/Ey$. If infinite quantities are available at the best quotes, then $\operatorname{dom} S_0^*$ is the product of the bid-ask intervals.

CSO duality

- Besides the above examples, one obtains e.g.
  - extensions to nonlinear and constrained models of financial markets as well as to Kabanov's currency market model.
  - extensions of the classical results of Rockafellar and Wets.
  - that in optimal control, the dual solutions are the subgradients of the cost-to-go functions.
  - ...
- While the above is almost a mechanical application of Rockafellar's conjugate duality, more involved arguments are required in establishing
  - existence of primal solutions and the absense of a duality gap (Ch 4).
  - existence of dual solutions (Ch 5).
- Conjugate duality can be applied in continuous-time models too. [Bismut, 1973] studied a special optimal control format in CD without establishing the existence of solutions.

Closedness criteria

- The above expressions for φ^{**} provide dual representations of the optimal value φ provided φ is proper and **lower semicontinuous** (lsc), i.e.
$$\liminf_{\nu\to\infty} \varphi(u^\nu) \ge \varphi(u)$$
whenever $u^\nu \to u$ in $\mathcal{U}$.
- The traditional "direct method" assumes that Ef is jointly lsc and $Ef(\cdot, u)$ is inf-compact uniformly in u.
- In e.g. financial models, the topological inf-compactness condition often fails but there is a measure theoretic counterpart (Komlós theorem) that works well in $\mathcal{N}$.

Closedness criteria

Theorem (Komlós)

If $(x^\nu)_{\nu=1}^\infty \subset L^0(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ *is* **almost surely bounded** *in the sense that*

$$\sup_\nu |x^\nu(\omega)| < \infty \quad P\text{-a.s.}$$

then there is a sequence of convex combinations $\bar{x}^\nu \in \mathrm{co}\{x^\mu \,|\, \mu \geq \nu\}$ *that converges almost surely in* L^0.

The following infinite-dimensional version of Theorem 8.4 from **Convex Analysis** gives a sufficient condition for the boundedness condition in the dynamic setting.

Theorem

Let $C : \Omega \rightrightarrows \mathbb{R}^n$ *be closed convex-valued and* $\mathcal{F}$*-measurable. If* $\{x \in \mathcal{N} \,|\, x \in C^\infty \ a.s.\} = \{0\}$*, then every sequence in* $\{x \in \mathcal{N} \,|\, x \in C \ a.s.\}$ *is almost surely bounded.*

The lower bound has been relaxed in [Perkiö, 2014]

Closedness criteria

Given $(\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $f(\bar{x}, \bar{u}, \omega) < \infty$, let

$$f^\infty(x, u, \omega) := \lim_{\alpha \nearrow \infty} \frac{f(\bar{x} + \alpha x, \bar{u} + \alpha u, \omega) - f(\bar{x}, \bar{u}, \omega)}{\alpha}.$$

Theorem

Assume that f is bounded from below and that

$$\{x \in \mathcal{N} \mid f^\infty(x(\omega), 0, \omega) \leq 0 \text{ a.s.}\}$$

is a linear space. Then

$$\varphi(u) = \inf_{x \in \mathcal{N}} Ef(x, u)$$

is $\sigma(\mathcal{U}, \mathcal{Y})$-lsc and the inf is attained for every $u \in \mathcal{U}$.

The lower bound has been relaxed by [Perkkiö, 2014].

Closedness criteria

Example (Optimal stopping)

When

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

we have $f^\infty = f$ and

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

so the linearity condition is always satisfied.

Closedness criteria

Example (Optimal investment)

When

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

we get

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

If v is nonconstant and $D_t(\omega)=\mathbb{R}^J$, the linearity condition becomes the **no-arbitrage** condition

$$x\in\mathcal{N}:\ \sum x_t\cdot\Delta S_{t+1}\geq 0 \implies \sum x_t\cdot\Delta S_{t+1}=0.$$

Example

With transaction costs, we get the **robust no-arbitrage** condition introduced by [Schachermayer, 2004].

Closedness criteria

The linearity condition may hold even under arbitrage.

Example

It holds if $S_t^\infty(x,\omega) > 0$ for $x \notin \mathbb{R}^J_-$.

Example

In [Çetin and Rogers, 2007] with

$$S_t(x,\omega) = x^0 + s_t(\omega)\psi(x^1)$$

one has $S_t^\infty(x,\omega) = x^0 + s_t(\omega)\psi^\infty(x^1)$. When $\inf\psi' = 0$ and $\sup\psi' = \infty$ we have $\psi^\infty = \delta_{\mathbb{R}_-}$, so the condition in Example 27 holds.

Example

If $S_t(\cdot,\omega) = s_t(\omega)\cdot x$ for a componentwise strictly positive price process s and $D_t^\infty(\omega) \subseteq \mathbb{R}^J_+$ (infinite short selling is prohibited) then linearity condition holds.

Summary

- Convex stochastic optimization unifies many models in stochastic control, operations research and mathematical finance.
- The minimalist structure yields simplifications and extensions of existing techniques e.g. on existence, duality, optimality conditions and numerics.
- A similar approach works also with continuous-time models but the mathematics gets more complicated/interesting:
    - path spaces?
    - stochastic integrals?
    - admissibility?