B Fenchel conjugation

Definition B.1 For a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ the *(Fenchel) conjugate* function of f is $f^* : \mathbb{R}^n \to [-\infty, +\infty]$, given by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} [\xi^t x - f(x)].$$

By repeating the conjugation operation one also defines the *(Fenchel) biconjugate* of f, which is simply given by $f^{**} := (f^*)^*$.

Example B.2 Consider $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(x) := \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ +\infty & \text{if } x < 0. \end{cases}$$

Observe that this function is convex. Then (counting $0 \log 0$ as 0) we clearly have $f^*(\xi) = \sup_{x \geq 0} \xi x - x \log x$ for the conjugate. For an *interior* maximum in $\mathbb{R}_+$ (by concavity of the function to be maximized) the necessary and sufficient condition is $\xi - \log x - 1 = 0$, i.e., $x = \exp(\xi - 1)$, which gives the value $\xi x - x \log x = \exp(\xi - 1)$. Since this value is positive, we conclude that the point $x = 0$ stands no chance for the maximum, i.e., the maximum is always interior, as calculated above, giving $f^*(\xi) = \exp(\xi - 1)$ for the conjugate function. We can also determine the biconjugate function: by definition, $f^{**}(x) = \sup_{\xi \in \mathbb{R}} x\xi - \exp(\xi - 1)$. If $x < 0$, then, by $\exp(\xi - 1) \to 0$ as $\xi \to -\infty$, the supremum value is clearly $+\infty$. Hence, $f^{**}(x) = +\infty$ for $x < 0$. If $x > 0$, then setting the derivative of the concave function $\xi \mapsto x\xi - \exp(\xi - 1)$ equal to zero gives a solution (whence a global maximum) for $\xi = \log x + 1$. Hence $f^{**}(x) = x \log x$ for $x > 0$. Finally, if $x = 0$, then the supremum of $-\exp(\xi - 1)$ is clearly the limit value 0. So $f^{**}(0) = 0$. We conclude that $f^{**} = f$ in this example. The Fenchel-Moreau theorem below will support this observation.

Exercise B.1 Determine for each of the following functions f the conjugate function f^* and verify also explicitly if $f = f^{**}$ holds.
a. $f(x) = ax^2 + bx + c$, $a \geq 0$,
b. $f(x) = |x| + |x - 1|$,
c. $f(x) = x^a/a$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ (here $a \geq 1$).
d. $f = \chi_B$, where B is the closed unit ball in $\mathbb{R}^n$.

Example B.3 Let K be a nonempty convex cone in $\mathbb{R}^n$ (recall that a *cone* (at zero) is a set such that $\alpha x \in K$ for every $\alpha > 0$ and $x \in K$; cf. Definition 2.5.1 in [1]). Let $f := \chi_K$. Then

$$f^*(\xi) = \sup_{x \in K} \xi^t x = \begin{cases} 0 & \text{if } \xi \in K^*, \\ +\infty & \text{otherwise.} \end{cases}$$

Recall here that K^*, the *polar cone* of K, is defined by $K^* := \{\xi \in \mathbb{R}^n : \xi^t x \leq 0 \text{ for all } x \in K\}$. Hence, we conclude that $(\chi_K)^* = \chi_{K^*}$.

Denote the closure of K by $\bar{K}$. We also observe that $\xi \in \partial \chi_{\bar{K}}(0)$ is equivalent to $\xi^t x \leq 0$ for all $x \in \bar{K}$, i.e., to $\xi^t x \leq 0$ for all $x \in K$, i.e., to $\xi \in K^*$.

Lecture 7: Convex Analysis and Fenchel-Moreau Theorem

The main tools in mathematical finance are from theory of stochastic processes because things are random. However, many objects are convex as well, e.g. collections of probability measures or trading strategies, utility functions, risk measures, etc.. Convex duality methods often lead to new insight, computational techniques and optimality conditions; for instance, pricing formulas for financial instruments and characterizations of different types of no-arbitrage conditions.

Convex sets

Let X be a real topological vector space, and X^* denote the topological (or algebraic) dual of X. Throughout, we assume that subsets and functionals are proper, i.e., $\emptyset \neq C \neq X$ and $-\infty < f \not\equiv \infty$.

Definition 1. Let $C \subset X$. We call C affine if

$$\lambda x + (1-\lambda)y \in C \ \forall x, y \in C, \ \lambda \in]-\infty, \infty[,$$

convex if

$$\lambda x + (1-\lambda)y \in C \ \forall x, y \in C, \ \lambda \in [0,1],$$

cone if

$$\lambda x \in C \ \forall x \in C, \ \lambda \in]0, \infty].$$

Recall that a set A is called algebraic open, if the sets $\{t : x + tv\}$ are open in $\mathbb{R}$ for every $x, v \in X$. In particular, open sets are algebraically open.

Theorem 1. *(Separating Hyperplane Theorem) Let A, B be convex subsets of X, A (algebraic) open. Then the following are equal:*

1. *A and B are disjoint.*
2. *There exists an affine set $\{x \in X : f(x) = c\}$, $f \in X^*$, $c \in \mathbb{R}$, such that $A \subset \{x \in X : f(x) < c\}$ and $B \subset \{x \in X : f(x) \geq c\}$.*

If in addition A and B are cones, we may take $c = 0$.

Proof. 2) $\implies$ 1) is obvious, 1) $\implies$ 2) exercise. □

On a locally convex X, Separating Hyperplane Theorem gives an useful characterization of closed convex sets:

$$A \subset X \text{ is closed and convex} \iff A = \bigcap_\alpha \{x \in X : f_\alpha(x) \geq \alpha, \ f_\alpha \in X^*, \ \alpha \in \mathbb{R}\}.$$

If X is also separable, then one can replace $\mathbb{R}$ with $\mathbb{N}$ above.

Remark 1. *Let X be a d-dimensional Hilbert space ($d = \#$ of base vectors). We say that $C \subset X$ is Chebyshev set if for every $x \in X$ there exists unique $\tilde{y} \in C$ such that $\tilde{y} = \arg\min_{y \in C} ||x - y||$. Consider the following statement:*

$$C \text{ is Cheyshev set} \iff C \text{ is closed and convex.}$$

If $d < \infty$, then it is an easy exercise to verify that the statement above is true. However, for $d = \infty$, this an open problem (convexity in the implication "$\Rightarrow$"). Lesson: Be careful when $d = \infty$.

Convex functionals

Definition 2. We call f affine on X if

$$\lambda f(x) + (1 - \lambda) f(y) = f(\lambda x + (1 - \lambda y)) \ \forall x, y \in X, \ \lambda \in] - \infty, \infty[,$$

convex if

$$\lambda f(x) + (1 - \lambda) f(y) \leq f(\lambda x + (1 - \lambda) y) \ \forall x, y \in X, \ \lambda \in [0, 1],$$

positively homogeneous if

$$f(\lambda x) = \lambda f(x) \ \forall x \in X, \lambda \in]0, \infty],$$

sub-additive if

$$f(x + y) \leq f(x) + f(y) \ \forall x, y \in X.$$

Remark 2. *If f is convex and positively homogeneous, then f is sub-additive. Positively homogeneous and sub-additive f is called sub-linear.*

Theorem 2. *(Hahn-Banach Theorem) Let Y be a subspace of X, and f linear functional on Y. If there exists a sub-linear functional g on X such that $f \leq g$ on Y (and g continuous at 0), then there exists $\tilde{f} \in X^*$ such that $\tilde{f} = f$ on Y and $f \leq g$ on X.*

Proof. This is proven at the basic course of functional analysis. □

Remark 3. *Separating Hyperplane Theorem (SHT) and Hahn-Banach Theorem (HBT) are equal in the following sense: If we assume that SHT is true, then HBT is true. Conversely, if we assume that HBT is true, then SHT is true.*

Convex conjugates

We make an additional assumption that X is Hausdorff and locally convex.

Definition 3. The epigraph of f is defined as

$$\operatorname{epi} f := \{(x, \alpha) \in X \times \mathbb{R} : f(x) \leq \alpha\}.$$

If $\operatorname{epi} f$ is closed, we say that f is lower semicontinuous (abbreviated l.s.c.).

Lemma 3. *If f is l.s.c. and convex, then, for every $x \in X$,*

$$f(x) = \sup_{a \leq f} a(x),$$

where the supremum is taken over all continuous affine functionals on X.

Proof. The idea of proof: "If a point does not belong to the epigraph, then there is an affine minorant in between."

Let $(x_0, \alpha_0) \in X \times \mathbb{R} \setminus \operatorname{epi} f$. By Separating Hyperplane Theorem, there exists $(x^*, c) \in X^* \times \mathbb{R}$ such that

$$x^*(x) + \alpha c > x^*(x_0) + \alpha_0 c, \ \forall (x, \alpha) \in \operatorname{epi} f.$$

If $c \neq 0$, it can be scaled to $c = 1$. Then $\inf_{(x,\alpha)\in \operatorname{epi} f}\{x^*(x) + \alpha\} - x^*(x)$ is an affine minorant of f whose epigraph does not contain (x_0, α_0). So, assume $c = 0$. Choose $x_1 \in X$ such that $f(x_1) < \infty$. Then $(x_1, f(x_1) - 1) \notin \operatorname{epi} f$, and, by Separating Hyperplane Theorem, there exists $(y^*, c') \in X^* \times \mathbb{R}$ such that

$$y^*(x) + \alpha c' > y^*(x_1) + (f(x_1) - 1)c', \ \forall (x, \alpha) \in \operatorname{epi} f.$$

Since $f(x_1) < \infty$, we have $c' \neq 0$, and, by scaling, we can assume $c' = 1$. Choosing

$$\delta > \frac{y^*(x_0) + \alpha_0 - \inf_{(x,\alpha)\in \operatorname{epi} f}\{y^*(x) + \alpha\}}{\inf_{(x,\alpha)\in \operatorname{epi} f}\{x^*(x) + \alpha\} - x^*(x_0)}$$

and setting $z^* = \delta x^* + y^*$ yields

$$\inf_{(x,\alpha)\in \operatorname{epi} f}\{z^*(x)+\alpha\} \geq \delta \inf_{(x,\alpha)\in \operatorname{epi} f}\{x^*(x)+\alpha\} + \inf_{(x,\alpha)\in \operatorname{epi} f}\{y^*(x)+\alpha\} > z^*(x_0)+\alpha_0.$$

So, $\inf_{(x,\alpha)\in \operatorname{epi} f}\{z^*(x) + \alpha\} - z^*(x)$ is an affine minorant of f whose epigraph does not contain (x_0, α_0). □

Definition 4. Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$. Then $f^* : X^* \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$f^*(x^*) := \sup_{x\in X}\{x^*(x) - f(x)\}$$

is called the convex conjugate of f, and the mapping

$$f \mapsto f^*$$

is called Legendre-Fenchel transformation.

Theorem 4. *(Fenchel-Moreau Theorem) If f is l.s.c. and convex, then Legendre-Fenchel transformation is bijection: $f^{**} = f$, where*

$$f^{**}(x) := \sup_{x^*\in X^*}\{x^*(x) - f^*(x^*)\}.$$

Proof. By the definition of f^*,

$$f(x) \geq \sup_{x^* \in X^*} \{x^*(x) - f^*(x)\}, \text{ for every } x \in X.$$

i.e. $f \geq f^{**}$. Let a be an affine minorant of f; $a \leq f$, so, $a^* \geq f^*$, so, $a^{**} \leq f^{**}$, but since a is affine, $a^{**} = a$. So, every affine minorant of f is an affine minorant of f^{**}. By Lemma 3, $f \leq f^{**}$. □

We could have proved the theorem also making the following observation: for fixed $x \in X$, the directional derivative

$$f'(x; y) := \lim_{\epsilon \downarrow 0} \frac{f(x + \epsilon y) - f(x)}{\epsilon}$$

is sub-linear as a functional of y. By Hahn-Banach Theorem, there exists $\tilde{f}' \in X^*$ such that

$$\tilde{f}'(y) \leq f'(x; y) \leq f(x + y) - f(x), \text{ for every } y \in X.$$

So, $f(x) + f^*(\tilde{f}') = \tilde{f}'(x)$, which completes the proof. We say that $\tilde{f}'$ is a sub-gradient of f at x, and denote $\tilde{f}' \in \partial f(x)$. The collection of all sub-gradients ∂f is called sub-differential.

References

These lecture notes were written in a hurry and may contain erors. See sources listed below for details and full treatment of the things presented and discussed on the lecture.[Che13],[Roc70],[RWW98],[Pen12]

References

[Che13] Patrick Cheridito. Convex analysis, lecture notes, 2013.

[Pen12] T. Pennanen. Introduction to convex optimization in financial markets. *Math. Program.*, 134(1, Ser. B):157–186, 2012.

[Roc70] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[RWW98] Ralph Tyrrell Rockafellar, Roger J.-B Wets, and Maria Wets. *Variational analysis.* Grundlehren der mathematischen Wissenschaften. Springer, Berlin, Heidelberg, New York, 1998. Autres tirages : 2004.

Proposition B.4 *Let $f,\ g : \mathbb{R}^n \to (-\infty, +\infty]$.*

(i) If $f \geq g$ then $f^* \leq g^*$.

(ii) If $f^*(x) = -\infty$ for some $x \in \mathbb{R}^n$, then $f \equiv +\infty$.

(iii) For every $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) \geq \xi^t x_0 - f(x_0) \ \textit{(Young's inequality).}$$

(iv) $f \geq f^{**}$.

(v) For every $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) = \xi^t x_0 - f(x_0) \text{ if and only if } \xi \in \partial f(x_0).$$

Exercise B.2 Give a proof of Proposition B.4.

Theorem B.5 (Fenchel-Moreau) *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be convex. Then*

$$f(x_0) = f^{**}(x_0) \text{ if and only if } f \text{ is lower semicontinuous at } x_0.$$

Proof. One implication is very simple: if $f(x_0) = f^{**}(x_0)$, and if $x_n \to x_0$ then $\liminf_n f(x_n) \geq \liminf_n f^{**}(x_n)$ by Proposition B.4(iv). Also, $\liminf_n f^{**}(x_n) \geq f^{**}(x_0)$ because every conjugate, being the supremum of a collection of continuous functions, is automatically lower semicontinuous. So we conclude that $\liminf_n f(x_n) \geq f^{**}(x_0) = f(x_0)$, i.e., f is lower semicontinuous at x_0.

In the converse direction, by Proposition B.4(iv) it is enough to prove $f^{**}(x_0) \geq r$ for an arbitrary $r < f(x_0)$, both when $f(x_0) < +\infty$ and when $f(x_0) = +\infty$.

Case 1: $f(x_0) < +\infty$. It is easy to check that $C := \text{epi } f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}$, the epigraph of f, is a convex set in $\mathbb{R}^{n+1}$ (this is Theorem 3.2.2 in [1] – as can be seen immediately from its proof, it continues to hold for functions with values in $(-\infty, +\infty]$ and we know already that this theorem also holds for sets with empty interior). Hence, the closure cl C is also convex. We claim now that $(x_0, r) \notin \text{cl } C$. For suppose (x_0, r) would be the limit of a sequence of points $(x_n, y_n) \in C$. Then $y_n \geq f(x_n)$ for each n, and in the limit this would give $r \geq \liminf_n f(x_n) \geq f(x_0)$ by lower semicontinuity of f at x_0. This contradiction proves that the claim holds. We may now apply separation [1, Theorem 2.4.10]: there exist $\alpha \in \mathbb{R}$ and $p =: (\xi_0, \mu) \neq (0, 0)$, with $\xi_0 \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, such that

$$\xi_0^t x + \mu y \leq \alpha < \xi_0^t x_0 + \mu r \text{ for all } (x, y) \in C. \tag{3}$$

It is clear that $\mu \leq 0$ by the definition of C. Also, it is obvious that $\mu \neq 0$ (just consider what happens if we take $(x, y) = (x_0, f(x_0))$ in (3) – and we may do this by virtue of $f(x_0) \in \mathbb{R}$). Hence, we can divide by $-\mu$ in (3) and get

$$\xi_1^t x - f(x) \leq \xi_1^t x_0 - r \text{ for all } x \in \text{dom } f.$$

Notice that this inequality continues to hold outside dom f as well; thus, $f^*(\xi_1) \leq \xi_1^t x_0 - r$, which implies the desired inequality $f^{**}(x_0) \geq r$.

Case 2a: $f \equiv +\infty$. In this case, the desired result is trivial, for $f^* \equiv -\infty$, so $f^{**} \equiv +\infty$.

Case 2b: $f(x_1) < +\infty$ for some $x_1 \in \mathbb{R}^n$. We can repeat the proof of Case 1 until (3). If μ happens to be nonzero, then of course we finish as in Case 1. However, if $\mu = 0$ we only get

$$\xi_0^t x \le \alpha < \xi_0^t x_0 \text{ for all } x \in \text{dom } f$$

from (3). We then repeat the full proof of Case 1, but with x_0 replaced by x_1 and r by $f(x_1) - 1$. This gives the existence of $\xi \in \mathbb{R}^n$ such that

$$\xi^t x - f(x) \le \xi^t x_1 - f(x_1) + 1 \text{ for all } x \in \text{dom } f.$$

Now for any $\lambda > 0$, observe that by the two previous inequalities

$$f(x) \ge (\xi + \lambda \xi_0)^t x - \xi^t x_1 + f(x_1) - 1 - \alpha\lambda \text{ for all } x \in \mathbb{R}^n,$$

which implies $f^*(\xi + \lambda\xi_0) \le \xi^t x_1 - f(x_1) + 1 + \lambda\alpha$. By definition of $f^{**}(x_0)$, this gives

$$f^{**}(x_0) \ge \lambda(\xi_0^t x_0 - \alpha) + \xi^t x_0 - \xi^t x_1 + f(x_1) - 1,$$

which implies $f^{**}(x_0) = +\infty$, by letting λ go to infinity (note that $\xi_0^t x_0 - \alpha > 0$ by the above). QED

Corollary B.6 (bipolar theorem for cones) *Let K be a closed convex cone in $\mathbb{R}^n$. Then $K = K^{**} := (K^*)^*$.*

PROOF. Observe that $f := \chi_K$ is a lower semicontinuous convex function. Hence, $f^{**} = f$ by Theorem B.5. By Example B.3 we know that $f^* = \chi_{K^*}$, so $f^{**} = \chi_{K^{**}}$ follows by another application of this fact. Hence $\chi_K = \chi_{K^{**}}$. QED

Exercise B.3 Prove Farkas' theorem (see Exercise 3.5) by means of Corollary B.6.

Exercise B.4 Redo Exercise 3.3 by making it a special case of Corollary B.6.

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* Wiley, New York, 1993.

[2] J. van Tiel, *Convex Analysis: An Introductory Text.* Wiley, 1984.