Section 12 (Conjugates of Convex Functions)

A closed cvx set in $\mathbb{R}^n$ is the intersection of the half spaces which contain it (Thm 11.5). Therefore, the epigraph of a clsd proper cvx fct on $\mathbb{R}^n$ is the intersection of the clsd half-spaces in $\mathbb{R}^{n+1}$ which contain it. We will determine the defn of conjugacy by translating this geometric idea to fcts.

Represent the hyper-planes in $\mathbb{R}^{n+1}$ by the linear functions

$$(x, \mu) \longrightarrow \langle x, b\rangle + \mu b_0$$

w/ $b \in \mathbb{R}^n$ & $\beta_0 = 0$ or $\beta_0 = -1$ (in general, $\beta_0 \in \mathbb{R}$ but as any scalar multiple of this linear fct will define the same hyperplane only considering $\beta_0 = 0$ and $\beta_0 = -1$ will suffice).

Hyperplanes for $\beta_0 = 0$ (called vertical) are

$$\{(x, \mu) \mid \langle x, b\rangle = \beta\}, \quad 0 \neq b \in \mathbb{R}^n, \quad \beta \in \mathbb{R}$$

and $\beta_0 = -1$ are

$$\{(x, \mu) \mid \langle x, b\rangle - \mu = \beta\}, \quad b \in \mathbb{R}^n, \quad \beta \in \mathbb{R},$$

which are graphs of affine fcts $h(x) = \langle x, b\rangle - \beta$ on $\mathbb{R}$.
Every closed half-space in $\mathbb{R}^{n+1}$ is thus one of the following types:

1. (vertical) $\{(x, \mu) \mid \langle x, \beta\rangle \le \beta\} = \{(x, \mu) \mid h(x) \le 0\}, \quad b \neq 0,$
2. (upper) $\{(x, \mu) \mid \mu \ge \langle x, b\rangle - \beta\} = \text{epi } h$
3. (lower) $\{(x, \mu) \mid \mu \le \langle x, b\rangle - \beta\}$

Theorem 12.1

A clsd cvx fct f is the pointwise supremum of the collection of all affine fcts h s.t. $h \le f$.

Proof

Assume f is proper (otherwise it is trivial since, $\text{cl} f = f$ & $\text{cl}\, g \equiv -\infty$ for any improper fct g, therefore $f \equiv -\infty$). epi f is a closed cvx set (Thm 7.1) and therefore epi f is the intersect of the upper and vertical h.s. in $\mathbb{R}^{n+1}$ containing epi f (note: no lower h.s. could contain epi f). The h.s. cannot all be vertical since f is assumed to be proper. Note that the upper closed h.s. are simply the epigraphs of the affine fcts h s.t. $h \le f$ and the intersection of epi h for all such h is just the pointwise supremum of the fcts $h \le f$. Therefore to prove the theorem we are only left to show that intersection of the upper & vertical h.s is identical to the intersection of simply the upper h.s.

Suppose

$$V = \{(x, \mu) \mid 0 \ge \langle x, b_1 \rangle - \beta_1 = h_1(x)\}$$

is a vertical h.s. and $(x_0, \mu_0) \notin V$. It is enough to show $\exists$ affine fct h s.t. $h \le f$ & $\mu_0 < h(x_0)$. There exists at least one affine fct h_2 s.t. $\text{epi} f \subset \text{epi} h_2$ i.e. $h_2 \le f$.

$\forall x \in \operatorname{dom} f$ $\quad h_1(x) \le 0$ & $h_2(x) \le f(x)$ and hence

$$\lambda h_1(x) + h_2(x) \le f(x) \qquad \forall \lambda \ge 0$$

(this inequality holds $\forall x$ since it is trivial if $x \notin \operatorname{dom} f$). Fix any $\lambda \ge 0$ & define h by

$$h(x) = \lambda h_1(x) + h_2(x) = \langle x, \lambda b_1 + b_2 \rangle - (\lambda \beta_1 + \beta_2)$$

and we have $h(x) \le f$. Since $(x_0, \mu_0) \notin V$ we know $h(x_1) > 0$, then $h(x_0) = \lambda h_1(x_0) + h_2(x_0)$ and by choosing large enough λ (specifically $\lambda > \max\{0, \frac{\mu_0 - h_2(x_0)}{h_1(x_0)}\}$) we ensure $h(x_0) > \mu_0$. $\square$

Corollary 12.1.1

If f is any fct from $\mathbb{R}^n$ to $[-\infty, \infty]$, then $\operatorname{cl}(\operatorname{conv} f)$ is the pointwise supremum of the collection of all affine fcts on $\mathbb{R}^n$ majorized by f.

Corollary 12.1.2

Given any proper cvx fct f on $\mathbb{R}^n$, there exists some $b \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ s.t. $f(x) \ge \langle x, b \rangle - \beta$ $\forall x$.

Definition

Let f be any closed cvx fct on $\mathbb{R}^n$. Consider

$$F^* = \{(x^*, \mu^*) \mid h(x) = \langle x, x^* \rangle - \mu^* \le f(x)\}.$$

Note that $h(x) \le f(x)$ $\forall x$ iff

$$\mu^* \ge \sup_{x \in \mathbb{R}^n} \langle x, x^* \rangle - f(x)$$

that is F^* is the epigraph of some fct f^* on $\mathbb{R}^n$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^n} \langle x, x^* \rangle - f(x) = -\inf_{x \in \mathbb{R}^n} f(x) - \langle x, x^* \rangle.$$

This f^* is the **conjugate** of f.

Remarks

- f^* is again a closed cvx fct. (Pointwise supremum of of the affine fcts $g(x^*) = \langle x, x^* \rangle - \mu$ s.t. $(x,\mu) \in \operatorname{epi} f$)
- conjugate of f^* is $f^{**} = f$. (f is pointwise supremum of the affine fcts $h(x) = \langle x, x^* \rangle - \mu^*$ s.t. $(x^*, \mu^*) \in \operatorname{epi} f$, i.e.,

$$f(x) = \sup_{x^* \in \mathbb{R}} \langle x, x^* \rangle - f^*(x^*) = -\inf_{x^* \in \mathbb{R}^n} f^*(x^*) - \langle x, x^* \rangle .)$$

- Corollary 12.1.1 allows us to define conjugate for any arbitrary $f : \mathbb{R}^n \longrightarrow [-\infty, \infty]$ since $f^* = g^*$ w/ $g = \operatorname{cl}(\operatorname{conv} f)$.

Examples

Consider the clsd proper cvx fct $f(x) = e^x$, $x \in \mathbb{R}^n$.

Then,

$$f^*(x^*) = \sup_{x \in \mathbb{R}} x x^* - e^x .$$

$\underline{x^* < 0}$ $x x^* - e^x$ can be made arbitrarily large by taking $x \longrightarrow -\infty$ so that $f^*(x^*) = \infty$

$\underline{x^* > 0}$ then $\frac{d}{dx}(x x^* - e^x) = x^* - e^x = 0 \Rightarrow x = \log(x^*)$. Therefore,

$f^*(x^*) = x^* \log(x^*) - x^*$

$\underline{x^* = 0}$ then $f^*(x^*) = \sup_{x \in \mathbb{R}^n} -e^x = 0$.

Therefore

$$f^*(x^*) = \sup_{x \in \mathbb{R}^n} x x^* - e^x = \begin{cases} x^* \log x^* - x^* & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \\ \infty & \text{if } x^* < 0 \end{cases}$$

Notice that

$$f^{**} = \sup_{x^* \in \mathbb{R}} x x^* - f^*(x^*) = \sup_{x^* \geq 0} x x^* - x^* \log x^* + x^* = e^x$$

That is, $f^{**} = f$ as expected. Notice that $\operatorname{dom} f = \mathbb{R}$ but $\operatorname{dom} f^* \neq \mathbb{R}$, this is studied further is §B.

Theorem 12.2

Let f be a cvx fct. The conjugate fct f^* is then a clsd cvx fct, proper iff f is proper. Moreover, $(\mathrm{cl}\, f)^* = f^*$ and $f^{**} = \mathrm{cl}\, f$.

Proof

Since f^* describes affine fcts majorized by f Cor. 12.1.1. gives $f^* = (\mathrm{cl}\, f)^*$. If f is improper then $\mathrm{cl}\, f = -\infty$ (by definition) and $f^* = (\mathrm{cl}\, f)^* = \infty$ which is improper similarly if f^* is improper this implies f is improper. The contrapositive completes the proof. $\square$

Corollary 12.2.1

The conjugacy operation $f \to f^*$ induces a symmetric one-to-one correspondence in the class of all clsd proper cvx fcts on $\mathbb{R}^n$.

Corollary 12.2.2

For any cvx fct f $\mathbb{R}^n$, one actually has

$$f^*(x^*) = \sup\{\langle x, x^*\rangle - f(x) \mid x \in \mathrm{ri}(\mathrm{dom}\, f)\}.$$

Example (Boyd and Vandenberghe)

Consider $f(x) = \frac{1}{2}\|x\|^2$ then,

$$(x^*)^T x - \frac{1}{2}\|x\|^2 \leq \|x^*\|_* \|x\| - \frac{1}{2}\|x\|^2 \leq \frac{1}{2}\|x^*\|^2.$$

The last inequality comes from the fact that $\|x^*\|_*\|x\| - \frac{1}{2}\|x\|^2$ is maximized at $x = x^*$.

Then, choose $x = \alpha x^*$ w/ $\alpha > 0$ so that $(x^*)^T x = \|x^*\|_* \|x\|$ and therefore $(x^*)^T x - \frac{1}{2}\|x\| = \frac{1}{2}\|x^*\|_*^2$ implying

$$f^*(x^*) = \sup_x\{(x^*)^T x - \frac{1}{2}\|x\|^2\} = \frac{1}{2}\|x^*\|_*^2$$

In the case of differentiable fcts we have an easier way to determine conjugates.

Legendre Transform (Rockafellar (§26) & Boyd/Vandenberghe (§ 3.3))

Defn

Let f be a real-valued fct on an open C on $\mathbb{R}^n$. The **Legendre conjugate** of the pair (C, f) is defined to be the pair (D, g) where D is the image of C under the gradient mapping ∇f, and g is the fct on D given by the formula

$$g(x^*) = \langle (\nabla f)^{-1}(x^*), x^* \rangle - f((\nabla f)^{-1}(x^*)).$$

or

$$g(x^*) = \langle x, x^* \rangle - f(x)$$

where x is s.t. $x^* = \nabla f(x)$

Note: $(C, f) \to (D, g)$ is the Legendre transformation.

Defn (Rockafellar §13)

A finite cvx fct f on $\mathbb{R}^n$ is said also to be co-finite if epi f contains no NON-vertical half-lines.

e.g. $f(x) = x^2$ is co-finite whereas $g(x) = e^x$ is NOT.

Note: (Corollary 8.5.2) f is co-finite if

$$(f0^+)(y) = \lim_{\lambda \to \infty} \frac{f(\lambda y)}{\lambda} = +\infty \qquad \forall\, y \neq 0.$$

(Lemma 26.7) for a differentiable cvx fct f on $\mathbb{R}^n$ to be co-finite it is nec. & suff. that

$$|\nabla f(x_i)| \to \infty \quad \forall \ \{x_i\} \ \text{s.t.} \ |x_i| \to \infty$$

Theorem 26.6

Let f be a finite differentiable cvx fct on $\mathbb{R}^n$. In order that ∇f be a one-to-one mapping from $\mathbb{R}^n$ onto itself, it is nec. & suff. that f be strictly cvx & co-finite. Then f^* is a differentiable strictly cvx co-finite fct & f^* is the same as the Legendre conjugate of f, i.e.

$$f^* = \langle (\nabla f)^{-1}(x^*), x^* \rangle - f((\nabla f)^{-1}(x^*)) \qquad \forall\ x^*$$

Note: The Legendre conjugate of f^* is then in turn f.

Example (Boyd and Vandenberghe)

Let $A \in \mathbb{R}^{n \times n}$ be nonsingular, $b \in \mathbb{R}^n$, f be a finite, differentiable, strictly cvx, and co-finite fct, and consider $g(x) = f(Ax+b)$. Then

$$g^*(x^*) = x^T x^* - g(x)$$

w/ x s.t. $x^* = \nabla g(x)$.

Therefore,

$$x^* = A^T \nabla f(Ax+b) \quad \Rightarrow \quad x = A^{-1} \nabla f^{-1}((A^{-1})^T x^*) - A^{-1} b.$$

Then

$$\begin{aligned} g^*(x^*) &= \left(A^{-1}(\nabla f)^{-1}((A^{-1})^T x^*)\right)^T x^* - b^T (A^{-1})^T x^* - f((\nabla f)^{-1}(A^{-1} x^*)) \\ &= (\nabla f)^{-1}((A^{-1})^T x^*)^T ((A^{-1})^T x^*) - f((\nabla f)^{-1}(A^{-1} x^*)) - b^T (A^{-1})^T x^* \\ &= f^*((A^{-1})^T x^*) - b^T (A^{-1})^T x^* \end{aligned}$$

Examples ($p, q \in \mathbb{R}$ are conjugates: $\frac{1}{p} + \frac{1}{q} = 1$)

	$f(x)$	$f^*(x^*)$
Rockafellar	$\frac{1}{p}\lvert x\rvert^p, \quad 1<p<\infty$	$\frac{1}{q}\lvert x^*\rvert^q, \quad 1<q<\infty$
	$\begin{cases} -\frac{1}{p}x^p & \text{if } x \geq 0, \quad 0<p<1 \\ \infty & \text{if } x<0 \end{cases}$	$\begin{cases} -\frac{1}{q}\lvert x^*\rvert^q & \text{if } x^*<0, \quad -\infty<q<0 \\ \infty & \text{if } x^* \geq 0 \end{cases}$
	$\begin{cases} -(a^2-x^2)^{1/2} & \text{if } \lvert x\rvert \leq a \\ \infty & \text{if } \lvert x\rvert > a \end{cases}$	$a(1+(x^*)^2)^{1/2}$
	$\begin{cases} -\frac{1}{2} - \log x & \text{if } x>0 \\ \infty & \text{if } x \leq 0 \end{cases}$	$\begin{cases} -\frac{1}{2} - \log(-x^*) & \text{if } x^*<0 \\ \infty & \text{if } x^* \geq 0 \end{cases}$
Vandenberghe and Boyd	$ax+b$	$\begin{cases} -b & \text{if } x^* = a \\ \infty & \text{if } x^* \neq a \end{cases}$
	$\begin{cases} \frac{1}{x} & \text{if } x>0 \\ \infty & \text{if } x \leq 0 \end{cases}$	$\begin{cases} \infty & \text{if } x^*>0 \\ -2(-x^*)^{1/2} & \text{if } x^* \leq 0 \end{cases}$
	$\frac{1}{2}x^T Q x$ w/ Q symmetric positive definite	$\frac{1}{2}(x^*)^T Q^{-1}(x^*)$
	Indicator fct on a cvx set $S \subseteq \mathbb{R}^n$ $I_S(x)$	$\underbrace{I_S^*(x^*) = \sup_{x \in S} (x^*)^T x}_{\text{Support fct of the set } S.}$
	$\lVert x\rVert$	$f^*(x^*) = \begin{cases} 0 & \lVert x^*\rVert \leq 1 \\ \infty & \text{otherwise} \end{cases}$