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 set 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

$$\langle x, \mu \rangle \rightarrow \langle x, b \rangle + \mu \beta_0$$

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

$$\{ (x, \mu) | \langle x, b \rangle = \beta^2, \quad 0 \neq b \in \mathbb{R}^n, \quad \beta \in \mathbb{R} \}$$

and $\beta_0=-1$ are

$$\{ (x, \mu) | \langle x, b \rangle - \mu = \beta^2, \quad b \in \mathbb{R}^n, \quad \beta \in \mathbb{R} \}$$

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) $\{ (x, \mu) | \langle x, b \rangle \leq \beta^2 = \{ (x, \mu) | h(x) \leq 0 \} \quad b \neq 0$

2. (upper) $\{ (x, \mu) | \mu \geq \langle x, b \rangle - \beta \}$ = epi $h$

3. (lower) $\{ (x, \mu) | \mu \leq \langle x, b \rangle - \beta \}$
Theorem 12.1

A closed convex funct $\mathcal{F}$ is the pointwise supremum of the collection of all affine functs $h$ s.t. $h \leq \mathcal{F}$.

Proof

Assume $\mathcal{F}$ is proper (otherwise it is trivial since, $\text{cl} \mathcal{F} = \mathcal{F}$ & $\text{cl} g = -\infty$ for any improper funct $g$, therefore $\mathcal{F} = -\infty$). $\text{epi} \mathcal{F}$ is a closed convex set (Thm 7.1) and therefore $\text{epi} \mathcal{F}$ is the intersection of the upper and vertical h.s. in $\mathbb{R}^{n+1}$ containing $\text{epi} f$ (note: no lower h.s. could contain $\text{epi} \mathcal{F}$). The h.s. cannot all be vertical since $\mathcal{F}$ is assumed to be proper. Note that the upper closed h.s. are simply the epigraphs of the affine functs $h$ s.t. $h \leq \mathcal{F}$ and the intersection of $\text{epi} h$ for all such $h$ is just the pointwise supremum of the functs $h \leq \mathcal{F}$. 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

$$V = \{(x, x) \mid 0 \leq \langle x, b \rangle - \beta, = h_1(x)\}$$

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

\[ \forall x \in \text{dom } f \quad h_1(x) \leq 0 \quad \& \quad h_2(x) \leq f(x) \quad \text{and hence} \quad \lambda h_1(x) + h_2(x) \leq f(x) \quad \forall \lambda \geq 0 \]

(this inequality holds if \( \exists \chi \) since it is trivial if \( \forall x \notin \text{dom } f \)). Fix any \( \lambda > 0 \) \& 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 \( h(x) \leq f \). Since \( (x_0, \mu_0) \notin V \) we know \( h(x_0) > 0 \) then \( h(x_0) = \lambda h_1(x_0) + h_2(x_0) \) and by choosing large enough \( \lambda \) (specifically \( \lambda > \max \{0, \frac{c_0 - h_2(x_0)}{h_1(x_0)}\} \)) we ensure \( h(x_0) > \mu_0 \).

\[ \square \]

**Corollary 12.1**

If \( f \) is any \( \text{fct} \) from \( \mathbb{R}^n \) to \( [-\infty, \infty] \), then \( \text{cl}(\text{conv } f) \)

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 \( \text{cvx} \) fct \( f \) on \( \mathbb{R}^n \) there exists

some \( b \in \mathbb{R}^n \) and \( B \in \mathbb{R}^n \) s.t. \( f(x) \geq \langle x, b \rangle - B \quad \forall \chi \).

**Definition**

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

\[ F^* = \{ (x^*, \mu^*) \mid h(x) = \langle x, x^* \rangle - \mu^* \leq f(x) \} \]

Note that \( h(x) \leq f(x) \quad \forall \chi \quad \text{iff} \quad \mu^* \geq \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 \textbf{conjugate} of \( f \).
Remarks

- $f^*$ is again a closed convex fct. (Pointwise supremum of the affine fcts $g(x) = \langle x, x^* \rangle - \mu$ s.t. $(x, \mu) \in \text{epi} F$)
- Conjugate of $f^*$ is $f^{**} = \overline{f}$. ($\overline{f}$ is pointwise supremum of the affine fcts $h(x) = \langle x, x^* \rangle - \mu^*$ s.t. $(x, \mu^*) \in \text{epi} F$, i.e.,
  \[ f(x) = \sup_{x \in \mathbb{R}} \langle x, x^* \rangle - f^{**}(x) = -\inf_{x \in \mathbb{R}} f^{**}(x) - \langle x, x^* \rangle. \])
- Corollary 12.1.1 allows us to define conjugate for any arbitrary $f : \mathbb{R}^n \to [0, \infty]$ since $f^* = g^*$ w/ $g = \text{cl}(\text{conv } f)$.

Examples

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

Then,

$\begin{align*}
\overset{x^* \in \mathbb{R}}{\sup} \ x^* - e^x
\end{align*}$

- $x^* < 0$ $\Rightarrow$ $x^* - e^x$ can be made arbitrarily large by taking $x \to -\infty$ so that $\overset{x \to -\infty}{\sup} f(x) = \infty$
- $x^* > 0$ then $\overset{x^*}{\sup} (\overset{x}{\sup} (x^* - e^x)) = x^* e^x = 0 \Rightarrow x = \log(x^*)$. Therefore,

$\begin{align*}
\overset{x^*}{\sup} (x^*) = x^* \log(x^*) - x^*
\end{align*}$

- $x^* = 0$ $\Rightarrow$ $\overset{x^*}{\sup} (x^*) = \overset{x^*}{\sup} - e^x = 0$.

Therefore

$\begin{align*}
\overset{x^*}{\sup} (x^*) = \overset{x^*}{\sup} x^* - e^x = \begin{cases} x^* \log x^* - x^* & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \\ \infty & \text{if } x^* < 0 \end{cases}
\end{align*}$

Notice that $\overset{x^*}{\sup} (x^*) = \overset{x^*}{\sup} x^* - \overset{x^*}{\sup} (x^*) = \overset{x^*}{\sup} x^* - \overset{x^*}{\sup} \log x^* + x^* = e^x$.

That is,

$f^{**} = \overline{f}$ as expected. Notice that $\text{dom } f = \mathbb{R}$ but $\text{dom } f^{**} = \mathbb{R}$.
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, $(cl f)^* = f^*$ and $f^{**} = cl f$.

Proof
Since $f^*$ describes affine fcts majorized by $f$ Cor. 12.1.1. gives $f^{**} = (cl f)^*$. If $f$ is improper then $cl f = -\infty$ (by definition) and $f^* = (cl f)^* = \infty$ 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 $f : \mathbb{R}^n \to \mathbb{R}$, one actually has
$$f^*(x^*) = \sup \{ \langle x, x^* \rangle - f(x) | x \in \text{ri}(\text{dom } f) \}.$$

Example (Boyd and Vandenberghe)
Consider $f(x) = \frac{1}{2} \|x\|^2$ then,
$$(x^*)^T x - \frac{1}{2} \|x\|^2 \leq \|x^*\|_\infty \|x\| - \frac{1}{2} \|x\|^2 \leq \frac{1}{2} \|x^*\|^2.$$

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

Then choose $x = \alpha x^*$ w/ $\alpha > 0$ so that $(x^*)^T x = \|x^*\|_\infty \|x\| \|x\|^2$ and therefore $(x^*)^T x = \frac{1}{2} \|x\|^2 = \frac{1}{2} \|x^*\|^2$ implying
$$f^*(x^*) = \sup \{ (x^*)^T x - \frac{1}{2} \|x\|^2 \} = \frac{1}{2} \|x^*\|^2.$$
In the case of differentiable fcts we have an easier way to determine conjugates.

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

**Defn.**

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

$$g(\mathbf{x}^*) = \langle (\nabla \mathcal{F})^+(\mathbf{x}^*), \mathbf{x}^* \rangle - \mathcal{F}( (\nabla \mathcal{F})^+(\mathbf{x}^*) )$$

or

$$g(\mathbf{x}^*) = \langle \mathbf{x}, \mathbf{x}^* \rangle - \mathcal{F}(\mathbf{x})$$

where $\mathbf{x}$ is s.t. $\mathbf{x}^* = \nabla \mathcal{F}(\mathbf{x})$

**Note:** $(\mathcal{C}, \mathcal{F}) \rightarrow (\mathcal{D}, g)$ is the **Legendre transformation**.

**Defn.** (Rockafellar §13)

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

* e.g. $\mathcal{F}(\mathbf{x}) = \mathbf{x}^2$ is co-finite whereas $\mathcal{G}(\mathbf{x}) = e^\mathbf{x}$ is NOT.

**Note:** (Corollary 8.5.2) $\mathcal{F}$ is co-finite if

$$\lim_{\lambda \to +\infty} \frac{\mathcal{F}(\lambda \mathbf{x})}{\lambda} = +\infty \quad \forall \mathbf{x} \neq 0.$$

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

$$|\nabla \mathcal{F}(\mathbf{x})| \to \infty \quad \forall \mathbf{x} \text{ s.t. } |\mathbf{x}| \to \infty$$
Let $f$ be a finite differentiable convex function on $\mathbb{R}^n$. In order that $Df$ be a one-to-one mapping from $\mathbb{R}^n$ onto itself, it is nec. & suff. that $f$ be strictly convex & co-finite. Then $f^*$ is a differentiable strictly convex co-finite function & $f^*$ is the same as the Legendre conjugate of $f$, i.e.

$$f^* = \langle (Df)^{-1}(x^*), x^* \rangle - f((Df)^{-1}(x^*)) + x^*$$

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

**Example (Boyd and Vandenberghe)**

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

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

subject to $x^* = Dg(x)$.

Therefore,$$
x^* = A^T Df(Ax + b) \quad \Rightarrow \quad x = A^T Df^{-1}(A^T x^*) - A^{-1} b.
$$Then

$$g^*(x^*) = (A^T (Df)^{-1}(A^T x^*))^T x^* - b(A^T x^*) - f((Df)^{-1}(A^T x^*))$$

$$= (Df)^{-1}(A^T x^*)^T ((A^T)^T x^*) - f((Df)^{-1}(A^T x^*)) - b^T (A^T)^T x^*$$

$$= f^*(A^T x^*) - b^T (A^T)^T x^*$$
Examples \( \text{\( p, q \in \mathbb{R} \) are conjugates}: \ \frac{1}{p} + \frac{1}{q} = 1 \)

| \( \frac{1}{p} |x|^p \), \( 1 < p < \infty \) | \( \frac{1}{q} |x|^q \), \( 1 < q < \infty \) |
|---|---|
| \( \frac{1}{p} \) \( |x|^p \) if \( x \geq 0 \), \( 0 < p \leq 1 \) | \( \frac{1}{q} \) \( |x|^q \) if \( x \geq 0 \), \( -\infty < q < 0 \) |
| \( \infty \) if \( x < 0 \) | \( \infty \) if \( x < 0 \) |

\( \frac{1}{2} (\alpha^2 - x^2)^{\frac{1}{2}} \) if \( |x| \leq \alpha \)

\( \infty \) if \( |x| > \alpha \)

<table>
<thead>
<tr>
<th>( \frac{1}{2} ) ( -\log x ) if ( x &gt; 0 )</th>
<th>( \frac{1}{2} ) ( -\log (-x^<em>) ) if ( x^</em> &lt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \infty ) if ( x \leq 0 )</td>
<td>( \infty ) if ( x^* \geq 0 )</td>
</tr>
</tbody>
</table>

\( \alpha x + b \)

| \( -b \) if \( x^* = \alpha \) | \( \infty \) if \( x^* = -\alpha \) |

| \( \frac{1}{2} x^T Q x \) | \( \frac{1}{2} (x^*)^T Q^* (x^*) \) |

\( Q \) symmetric positive definite

B. & V: Indicator fct on a cvx set \( S \subseteq \mathbb{R}^n \)

\( I_S(x) \)

| \( ||x|| \) | \( f^*(x^*) \) |
|---|---|
| 0 if \( ||x^*|| \leq 1 \) | \( \infty \) otherwise |

\( \text{Support fct of the set} \ S. \)