Conjugate Functions

Tuesday, February 2, 2021 9:38 PM

aka Fenchel – Legendre conjugate*

*when speaking, I sometimes say "dual" when I mean "conjugate", and vice-versa. They are distinct, though related... Sorry in advance.

or, Fenchel – Legendre Transform, which reduces to the Legendre – Transform when you're differentiable.

Def The F.-L.- conjugate of f is

$$f^*(y) = \sup_x \underbrace{\langle y, x \rangle}_{} - f(x)$$

BV'04 says $y^T x$ but that's just specializing to Eucl. space
For, e.g., matrices, use $\mathrm{tr}(Y^T X)$

Prop f^* is convex (whether f is or not)

proof $y \mapsto \langle y, x \rangle - f(x)$ is convex $\forall y$, and arbitrary supremem preserve convexity □

When f is differentiable and full domain, the supremum occurs when $\nabla_x(\langle y,x\rangle - f(x)) = 0$

i.e., $y = \nabla f(x)$, so $x^* = (\nabla f)^{-1}(y)$

$$\left.\begin{aligned} f^*(y) &= \langle y, x^* \rangle - f(x^*) \\ &= \langle \nabla f(x^*), x^* \rangle - f(x^*) \quad \text{w/ } x^* = (\nabla f)^{-1}(y) \end{aligned}\right\} \text{Legendre Transform}$$

Legendre Transform in 1D ... to give us intuition.

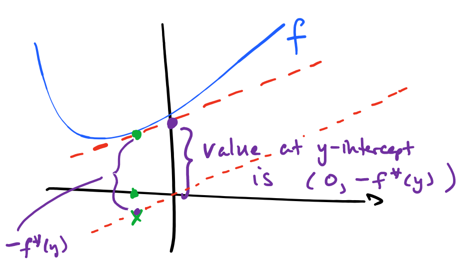
Assume f is strictly convex (so f' is strictly monotone, i.e., invertible)

$f^*(y) = \sup_x x \cdot y - f(x)$, maximized where $0 = y - f'(x)$, $f'(x) = y$

$x = (f')^{-1}(y)$

interpret as slope

a point x that has slope y



What is $f^*(\frac{1}{2}) = ?$

$y = \frac{1}{2}$ is the slope. ① Find the point x that has this slope (will be unique if f strictly convex)

slope = ½

② Now that we have x, evaluate $x \cdot y - f(x)$

Equation for red line: $\overset{\swarrow \text{slope}}{m} \cdot x + \overset{\swarrow \text{intercept}}{b}$

also, $\text{slope} \cdot (x - x_0) + f(x_0)$ } (algebra notation) where slope $= y$, $x_0 = x$

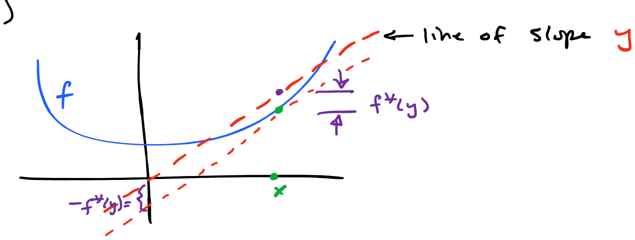
... so, intercept $= f(x_0) - \text{slope} \cdot x_0$

$= f(x) - yx$

$= -f^*(y)$

or... think of as finding the point x to maximize the (signed) separation of $\langle y,x\rangle$ and $f(x)$

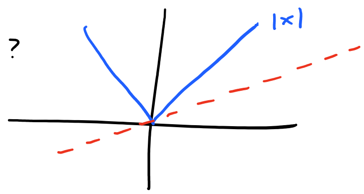
(want $\langle y,x\rangle$ on top of $f(x)$)



Ex: $f(x) = |x|$

What is $f^*(\frac{1}{2})$?

A:

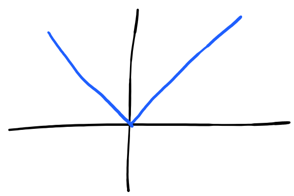


The x to make $\langle y,x\rangle - f(x)$ biggest (in this case, least negative) is at $x=0$. Gap is 0.

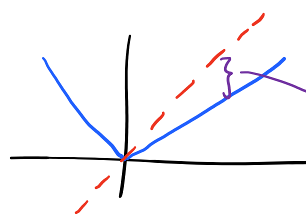
So... $f^*(\frac{1}{2}) = 0$

$\boxed{Q1}$ What is $f^*(2)$?

still $f(x) = |x|$



$\boxed{A1}$ It is $+\infty$



which point x makes this gap largest?

$x \to +\infty$

RULES

~~No running!~~

~~No horseplay!~~

~~No children under 5~~

Affine transformations: let $g(x) = f(Ax+b)$, assuming A invertible

$$g^*(y) := \sup_x \langle y,x\rangle - f(Ax+b), \quad \text{let } z = Ax+b \text{ or } x = A^{-1}(z-b)$$

$$= \sup_z \langle y, A^{-1}(z-b)\rangle - f(z)$$

$$= -\langle y,b\rangle + \sup_z \langle A^{-*}y, z\rangle - f(z)$$

$$= \boxed{\langle -y,b\rangle + f^*(A^{-*}y)} \quad \text{and} \quad \mathrm{dom}(g^*) = A^{T}\,\mathrm{dom}(f^*)$$

Sums of functions

let $f(x) = f_1(x) + f_2(x)$.

Is $f^*(x) = f_1^*(x) + f_2^*(x)$? NO

But... if "independent" (i.e., separable), in the sense (w/ $x = \begin{bmatrix} u \\ v \end{bmatrix}$, $u \in \mathbb{R}^{n_1}, v \in \mathbb{R}^{n_2}$, $n_1 + n_2 = n$)

if $f(u,v) = f_1(u) + f_2(v)$

then $f^*(w, z) = f_1^*(w) + f_2^*(z)$

Ex. Indicator Function of a set, $f(x) = I_C$

$$f^*(y) = \sup_x \langle x, y\rangle - I_C(x) = \sup_{x \in C} \langle x, y \rangle$$

This is called the support function of the set C

Q2 Let $C = \{x : \|x\|_p \le 1\}$, what is this set's support function

i.e., if $f(x) = I_C(x)$, what is f^* ?

A2 $f^*(y) := \sup_{\|x\|_p \le 1} \langle x, y \rangle$

By Hölder's ineq., $\langle x, y\rangle \le \|x\|_p \cdot \|y\|_q$ $\quad q = \frac{1}{1 - 1/p}$ i.e. $\frac{1}{p} + \frac{1}{q} = 1$

$\le 1 \cdot \|y\|_q$

so $f^*(y) \le \|y\|_q$

Also, there is an x to always make Hölder's tight

i.e., $\exists x$ s.t. $\langle x, y\rangle = \|x\|_p \cdot \|y\|_q$

so $\boxed{f^*(y) = \|y\|_q}$ the dual norm!

Ex: $p = q = 2$, choose $x = y / \|y\|_2$

So $\|x\|_2 = 1$

$\langle x, y\rangle = \frac{1}{\|y\|_2} \langle y, y\rangle$

$= \frac{1}{\|y\|_2} \|y\|_2^2 = \|y\|_2$

or $p = 1, q = \infty$

$\|x\|_1 \le 1$ choose x_i all 0 exept $\text{sign}(y_i)$ for $|y_i| = \|y\|_\infty$, so $\langle x, y\rangle = \text{sign}(y_i) \cdot y_i$

$= |y_i| = \|y\|_\infty$

or $p = \infty, q = 1$

$\|x\|_\infty \le 1$, choose all $x_i = \text{sign}(y_i)$ so $\langle x, y \rangle = \sum_i \text{sign}(y_i) \cdot y_i = \sum_i |y_i| = \|y\|_1$

So, the conjugate of the indicator function of a norm ball is the dual norm

$C = \{ \|x\| \le 1\}$, $f(x) = I_C \Rightarrow f^*(y) = \|y\|_*$ dual norm defined

$\|y\|_* = \sup_{\|x\| \le 1} \langle x, y \rangle$

What about dual of $g(x) = \|x\|$?

$$g^*(y) = \underbrace{\sup_x \langle x,y\rangle - \|x\|}_{\langle x,y\rangle \le \|x\|\cdot\|y\|_*}$$

like generic Hölder

so $g^*(y) \le \underbrace{\sup_x \|x\|\cdot\|y\|_* - \|x\|}$

if $\|y\|_* \le 1$, is maximized at $x=0$.

if $\|y\|_* > 1$, ... is maximized as $\|x\| \to \infty$

$$= \begin{cases} 0 & \|y\|_* \le 1 \\ +\infty & \|y\|_* > 1 \end{cases}$$

So, $g^*(y) = I_C(y)$, $C = \{x : \|x\|_* \le 1\}$

Converse to what we saw!

So, since (fact) $(\|\cdot\|_*)_* = \|\cdot\|$,

if $f(x) = \|x\|$, then $f^*(y) = I_{\{y:\|y\|_* \le 1\}}$, and $f^{**}(z) = \|z\|$

i.e., $\boxed{f = f^{**}}$ in this case

in general, true if f is convex and "nice"

BTW, is it possible for $f = f^*$? yes, for exactly 1 function,

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

We'll revisit the conjugate when we cover non-convex optimization