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

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

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 supremum preserve convexity $\square$

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.

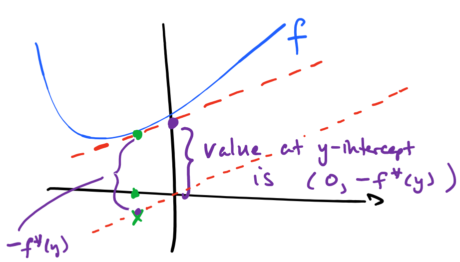
Assume f is strictly convex (so f' is strictly monotone, i.e., invertible)

$f^*(y) = \sup_x x \cdot y - f(x)$, 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 $f^*(1/2) = ?$

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

slope $= 1/2$

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

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

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

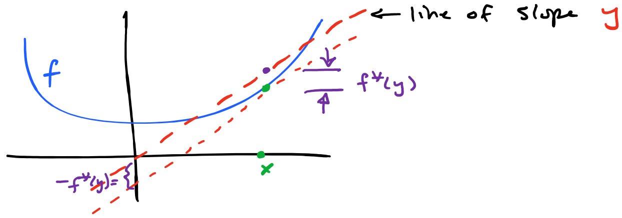
... so, intercept $= f(x_0) - \text{slope} \cdot x_0$

$= f(x) - yx$

$= -f^*(y)$

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

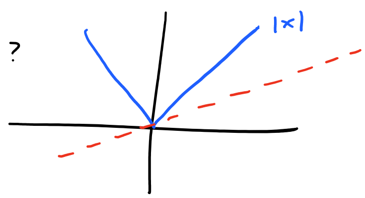
(want $\langle y, x\rangle$ on top of $f(x)$)



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

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

A:

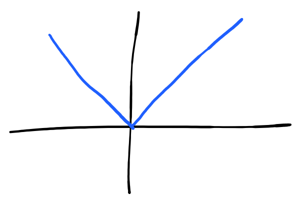


The x to make $\langle y, x\rangle - f(x)$ biggest (in this case, least negative) is at $x = 0$. Gap is 0.

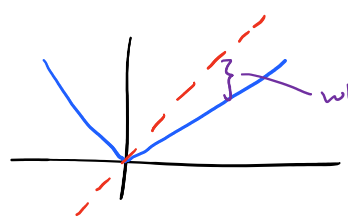
So... $f^*(\frac{1}{2}) = 0$

$\boxed{Q1}$ 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, $f(x) = \frac{1}{2}\|x\|^2$

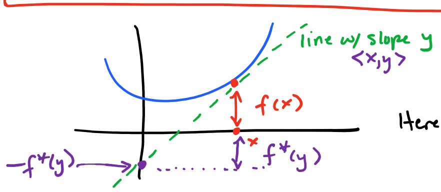
Properties

Fenchel-Young inequality (or just "Young's ineq." if differentiable)

Recall $f^*(y) := \sup_x \ \langle x,y\rangle - f(x)$ ... so $\forall x$, $f^*(y) \ge \langle x,y\rangle - f(x)$

$$\boxed{\forall x,y \quad f(x) + f^*(y) \ge \langle x,y\rangle}$$ F.Y. ineq.

Starts to hint at symmetry



Here, $\langle x,y\rangle = f(x) + f^*(y)$ since $y = f'(x)$

Corollary of F.Y. If $f \in \Gamma_0(\mathbb{R}^n)$ then (Thm 16.23 Bauschke + Combettes '11)

$$\big(y \in \partial f(x)\big) \Longleftrightarrow \big(f(x) + f^*(y) = \langle x,y\rangle\big)$$ equality

(why? $f^*(y) = \sup \langle x,y\rangle - f(x) = -\underbrace{\inf \ f(x) - \langle x,y\rangle}_{\text{Fermat's rule}}$

$0 \in \partial f(x) - y$)

but note you can show $f^* \in \Gamma_0(\mathbb{R}^n)$ under mild conditions, so then by symmetry

$$\big(x \in \partial f^*(y)\big) \iff \big(f(x) + f^*(y) = \langle x, y\rangle\big)$$

$$\big(y \in \partial f(x)\big) \iff \big(x \in \partial f^*(y)\big)$$

i.e. Corollary 16.24 $f \in \Gamma_0(\mathbb{R}^n) \Rightarrow (\partial f)^{-1} = \partial f^*$

So... finding $0 \in \partial f(x)$ Fermat's rule

is just finding $\boxed{\partial f^*(0)}$.

Convex Relaxations

A fact is that if $f: \mathbb{R}^n \to (-\infty, \infty]$ is proper, then

$(f$ is lsc and convex$) \iff (f = f^{**})$ Thm 13.32 Bauschke & Combettes '11

and if $f, g \in \Gamma_0(\mathbb{R}^n)$, $f = g^* \iff g = f^*$

so what about if f isn't convex?

Idea: f^* is convex even when f isn't

... so f^{**} is convex even when f isn't.

Also, $f \le g \Rightarrow f^* \ge g^*$

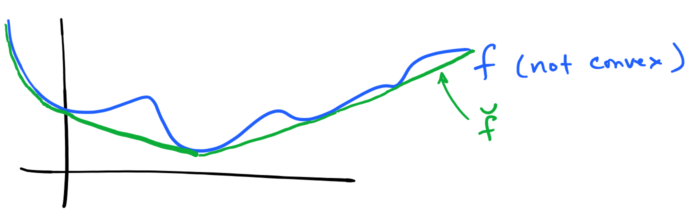
and so $\boxed{f^{**} \le f}$ and f^{**} is convex.

Def The "lsc convex envelope" $\boxed{\breve{f}}$ (informally, "convex relaxation") of f

is $\breve{f} = \sup\{ g \in \Gamma(\mathbb{R}^n) : g \le f \}$

and you can show

$$\text{epi}(\breve{f}) = \overline{\text{conv}(\text{epi}(f))}$$



Theorem (Becker, 2021) Convex optimization is NP-Hard

or...

all non-convex problems can be solved via convex optimization.

proof

$$\min \underbrace{f(x)}_{\text{non-convex}} = \min \underbrace{\breve{f}(x)}_{\text{convex}}$$

Corollary A convex problem is tractable only if we make certain structural assumptions, examples:

① we can evaluate $f(x)$

and/or ② we can evaluate $\partial f(x)$

etc.

The issue is that $\breve{f}(x)$ is too abstract ... we can't evaluate it.

But some help ...

Thm Usually $\breve{f} = f^{**}$. (i.e., need f proper, must assume $f^* \not\equiv +\infty$)

Common trick / heuristic

let $f = \underset{\uparrow}{h} + \underset{\nwarrow}{g}$ be nonconvex. We'd like to minimize f^{**} $(= \breve{f})$

nonconvex convex

... but $f^{**}(x)$ isn't easy to evaluate.

Instead, try solving $\min (h^{**} + g)$ since h^{**} is convex.

This isn't the "real"/"full" convex relaxation since

$$(h+g)^{**} \neq h^{**} + g^{**} \quad (\text{and } g^{**} = g)$$

... but it's still useful sometimes.

Example Compressed Sensing

(P) minimize $\|x\|_0$ → = #nonzero entries (very nonconvex)

s.t. $\|x\|_\infty \leq 1$, $Ax = b$ } NP-Hard

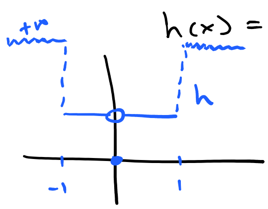
let $\underbrace{h(x) = \|x\|_0 + I_{\{\|x\|_\infty \leq 1\}}(x)}_{\text{this is separable ...}}$ and $g(x) = I_{\{Ax=b\}}(x)$

convex but not separable

mixes components together

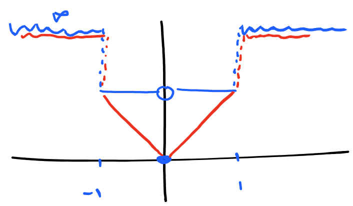
so only need treat 1D case

$$h(x) = \begin{Bmatrix} 0 & x = 0 \\ 1 & x \neq 0 \end{Bmatrix} + \begin{Bmatrix} 0 & |x| \leq 1 \\ \infty & |x| > 1 \end{Bmatrix}$$



$\boxed{Q3}$ What is h^{**}?

$\boxed{A3}$



$$h^{**}(x) = |x| + \begin{cases} 0 & |x| \le 1 \\ \infty & |x| > 1 \end{cases}$$

So our heuristic for (P) is $\min_x \|x\|_1$

s.t. $\|x\|_\infty \le 1$, $Ax = b$

often drop

"ℓ_1 is convex relaxation of ℓ_0"

Thm Compressed sensing (Candès, Romberg, Tao '04)

Drop ℓ_∞ constraints. Let sol'n to (P) be K, then

if A is iid Gaussian, $m \times n$, with $m = O(K \cdot \log(n/K))$

then ℓ_1 variant recovers the same argmin. (w/ high probability)