In preparation for a Math lecture:

**Subgradients + Subdifferentiability of Convex Functions**

Oddly, not in BV.

Assume $f$ is convex + proper: \( \exists x \in \mathbb{R}^n, f(x) < +\infty \)

\[ f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}, \quad \forall x, \quad f(x) > -\infty. \]

Define \( y^* \in \mathbb{R}^n \) as a subgradient of \( f \) at \( x \)

\[
f(x + z) \geq f(x) + y^*z \quad \forall z \in \mathbb{R}^n
\]

But...

If \( f \) is differentiable at \( x \), it is not

**n=1:** \( y \) is the slope of a line passing through \( (x, f(x)) \) and lying underneath the graph of \( f \).

\[ \left[ \begin{array}{c} y \\ -1 \end{array} \right] \]

**n \geq 1:** \[ \left[ \begin{array}{c} y \\ -1 \end{array} \right] \] is normal to a hyperplane in \( \mathbb{R}^{n+1} \) passing through \( \left[ \begin{array}{c} x \\ f(x) \end{array} \right] \) and lying below the graph of \( f \).

The set of all subgradients of \( f \) at \( x \) is denoted \( \partial f(x) \), the subdifferential of \( f \) at \( x \).

E.g. \( f(x) = |x| \), \( \partial f(0) = [-1, 1] \)
If $f$ is differentiable at $x$, then
\[ d_f(x) = \mathbb{E} \nabla f(x) \mathbb{I}. \]
In fact this is $I F F$
\[ f \text{ is convex}. \]

Note that $d_f(x)$ is always a closed, convex, non-empty, compact set.

E.g. $f(x) = \max_{1 \leq i \leq n} f(x_i)$ ($= x^n$ in BV notation)

What's $d_f(x)$ for $x = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 3 \end{bmatrix}$? Need
\[
\max \left( \begin{bmatrix} 1 + z_1 \\ 3 + z_2 \\ 2 + z_3 \\ 2 + z_4 \\ 3 + z_5 \end{bmatrix} \right) \geq 3 + y^T z \quad \forall z \in \mathbb{R}^n
\]

Clearly $e_1 \notin d_f(x)$ as RHS is $3 + z_1$ (take $z = e_1$)

$e_2 \in d_f(x)$ as RHS is $3 + z_2$.

In fact $d_f(x) = \text{conv}(e_2, e_5) = \left\{ \begin{bmatrix} 0 \\ c \\ c \\ 0 \end{bmatrix} : c \in [0, 1] \right\}$

Does this remind you of something?

Answer: (Fenchel) conjugate.

FHM (Fenchel-Hall-Yang)
\[ f(x) + f^*(y) \geq x^T y \]
with equality $I F F$ $y \in d_f(x)$.

Pf: Exercise in Borwein & Lewis.
Relationship to Directional Derivative

\[ f'(x; d) = \lim_{t \to 0} \frac{f(x+td) - f(x)}{t} \]

Then \( y \in \text{dom } f \) if and only if \( f'(x; d) \) exists for all \( d \in \mathbb{R}^n \).

Proof: By the theorem.

Chain Rule: simplest version.


Let \( f: \mathbb{R}^n \to \mathbb{R} \) convex, \( d \neq 0 \in \mathbb{R}^n \).

Let \( A \in \mathbb{R}^{n \times m} \), \( b \in \mathbb{R}^m \).

Let \( h \) be the convex function on \( \mathbb{R}^m \) defined by

\[ h(\xi) = f(A\xi + b) \quad \xi \in \mathbb{R}^m \]

Then \( \nabla h(\xi) = A^T df(A\xi + b) \)

means

\[ \{ A^T y; y \in \text{dom } f(A\xi + b) \} \]

Works even if \( A \) does not have full rank, e.g., \( A = 0 \).

Optimality Condition

\( 0 \in \text{dom } f \) if and only if \( x \) is a global minimizer of \( f \).

Proof: Immediate from definition.