Intuition of inequality in convex optimization:

① subgradient of f at x: $f(y) \geq f(x) + \langle g, y-x \rangle$

$y, x \in \mathbb{R} \Rightarrow$ $\frac{f(y)-f(x)}{y-x} \geq g \Rightarrow$
$y \neq x$

$\tan\theta = g$

No doubt that for f convex,
$\forall\, x \in \mathrm{dom} f$ has subgradient.

② $\partial f_1(x) + \partial f_2(x) \subseteq \partial(f_1+f_2)(x)$

$A^T\{\partial f[A(x)+b]\} \subseteq \partial h(x)$, $h(x) = f[A(x)+b]$

$\mathrm{conv}\left(\bigcup_{i \in I(x)} \partial f_i(x)\right) \subseteq \partial f(x)$, $f(x) = \max_{i \in I} f_i(x)$

Abstractly, doing subgradient then do operations will make sets smaller.

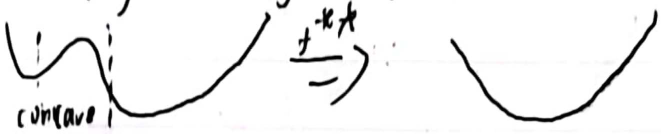
③ $f^*(y) = \max_x \{\langle y, x\rangle - f(x)\}$ is conjungate function,

of $f(x)$. If take x as displacement, y as force, $y^T x$ is external force on the system, $f(x)$ can be regarded as potential energy, $f^*(y)$ is other energy kinds produced by y.

④ Fenchel's inequ.: $f(x) + f^*(y) \geq \langle y, x \rangle$

Energy of the system has an U.B.
which is the ~~energy~~ work done by y on system.

⑤ $f^{**} \le f$: f^{**} must be convex. If f not, consider the concave region, they must also be convex after transformation:



⑥ conjugate subgradient theorem: f: proper, convex

i) $\langle x, y\rangle = f(x) + f^*(y)$ $\Leftrightarrow$ ii) $y \in \partial f(x)$ $\Leftrightarrow$ iii) For f closed, $x \in \partial f^*(y)$

consider $F(x,y) = f(x) + f^*(y) - \langle x, y \rangle \geq 0$ by Fenchel's ineq.

$\Rightarrow$ $F(x,y)=0$ is global minimal of a convex function $\Rightarrow$

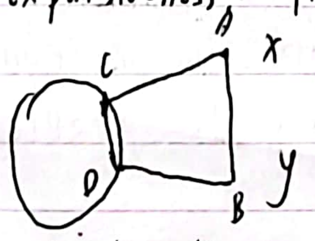
$$\begin{cases} \partial_x F(x,y) = \partial f(x) - y \ni 0 \\ \partial_y F(x,y) = \partial f^*(y) - x \ni 0 \end{cases} \Rightarrow \begin{matrix} y \in \partial f(x) \\ x \in \partial f^*(y) \end{matrix}$$

Closeness of f is for $f^{**} = f$, otherwise theorem may not hold to f^*, see proof for more detail.

⑦ nonexpansiveness: $\| \mathrm{prox}_f(x) - \mathrm{prox}_f(y) \| \le \| x - y \|$

$$\downarrow$$

$$|CD| \le |AB|$$



The validation of C, D in feasible set is "prox" is an operation that makes points x, y to the "nearest" points on feasible set.

⑧ f is L-smooth over convex set: $f(y) \le f(x) + \langle \nabla f(x), y-x \rangle + \frac{L}{2}\|x-y\|^2$

For $f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}(y-x)^T \nabla^2 f(\xi)(y-x)$, $\xi = x + \theta(y-x)$, $\theta \in (0,1)$

$\nabla^2 f(\xi) = \frac{|\nabla f(y) - \nabla f(x)|}{\|y-x\|} \le L \Rightarrow f(y) \ldots$

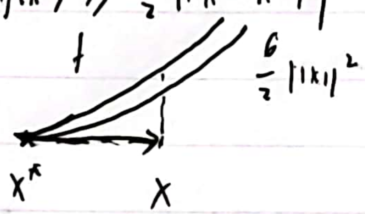
This is also U.B for estimating $f(y)$, $f(y) \ge f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|$ is L.B. This ineq. is less intuision,

⑨ $\langle \nabla f(x) - \nabla f(y), x-y \rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$ is for using gradient to measure distance. For some goals that is easy to measure angle but not good at distance may help. (joking)

⑩ Strongly-convex: $f(x) - \frac{\sigma}{2}\|x\|^2$ is convex $\Rightarrow$ more "convex" than $\frac{\sigma}{2}\|x\|^2$, Funtion graphing like

is not strongly-convex.

⑪ For strongly-convex function f: $f(x) - f(x^*) \ge \frac{\sigma}{2}\|x - x^*\|^2$, x^* is unique min of f:



(Basically like this, but graph is for $f(x) - f(x^*) \ge \frac{\sigma}{2}\|x\|^2 - \frac{\sigma}{2}\|x^*\|^2$

⑫ $T_L(x) = \text{prox}_{\frac{g}{L}}[x - \frac{1}{L}\nabla f(x)]$

$\downarrow$ projection $\qquad \downarrow$ gradient descent

If gradient descent makes a point ~~T~~ out of feasible set, the method will be meaningless.

13 Fundamental prox-grad ineq: $F = f + g$

$F(x) - F(T_L(y)) \geq \frac{L}{2}\|x - T_L(y)\|^2 - \frac{L}{2}\|x-y\|^2 + \ell_f(x,y)$

$\ell_f(x,y)$: $f(x) - f(y) - \langle \nabla f(y), x-y \rangle$

want to get x from y, $\ell_f(x,y)$ is gradient descent.

In the end, the most pleasant thing is we can take $x \in \mathrm{dom} f$ where f is not differentiable to be $\partial f(x)$ to estimate the optimal point or stationary point, which is we can do " derivative" on non-differential points.