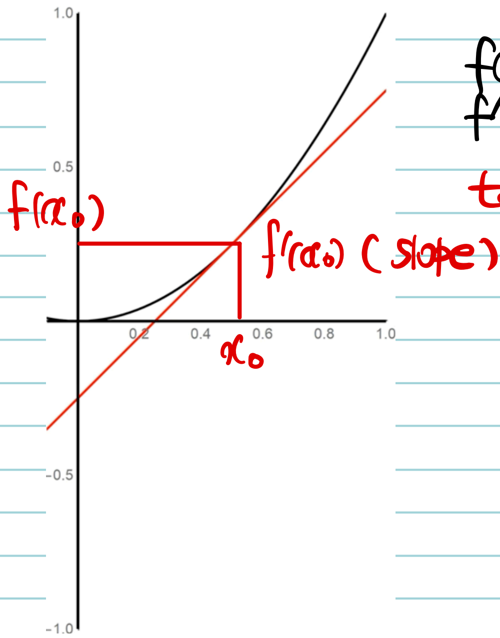
Lengendre Fenchel Transform (Convex Conjugate)

: Transforming certain function to new functions while retaining all of the information contained.

$f(x)$ $\qquad$ $f'(x) = p$ $\quad \rightarrow$ We want to find a new function $g(p)$, contains all of the information $f(x)$
$\Leftrightarrow$ possible to recover $f(x)$ from $g(p)$

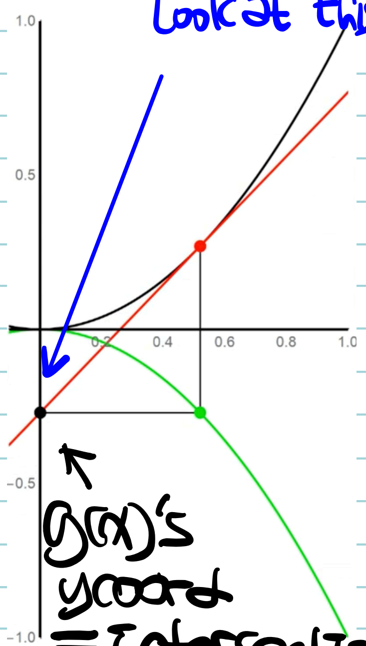
$f(x) = x^2$



$f(x) = x^2$
$f'(x) = p = 2x$

tangent line : $t(x, x_0) = f(x_0) + (x - x_0) f'(x_0)$

Look at this part!



We can now compose a function x coordinate of red dot and y coordinate of black dot.
$\rightarrow$ Green function in the end and this function is $g(x)$

g(x)'s ycoord = intersection of tangent of $f'(x)$

$g(x) = t(0, x) = f(x) - x f'(x)$
$= f(x) - xp \quad = x^2 - 2x^2 = -x^2$

- But we are not done yet. B/c Legendre f is a function of different variable than original function
- It is a function of derivative of original function(p).
- Therefore, we should Express x as a function of P, and put that into our expression for g.

$$f'(x) = p = 2x \Rightarrow x = \frac{p}{2}$$

$$x(p) = (f')^{-1}(P)$$ [compute inverse of $f'(x)$]

$$g(p) = g(x(p)) = -p^2/4$$

$$g(p) = f(x(p)) - x(p)p$$
$$g(p) = x(p)p - f(x(p))$$

When does Legendre transformation fails?

- A geometric construction is possible for any function
- But it doesn't actually yield the final Legendre transform as that is a function g of p. (not of x)
- To replace argument x with p, we have to find inverse of p of x
- If first derivative of f is not invertable, than f has no Legendre form.
- Thus, $f'(x)$ have to be a monotonic function

Mathematical definition

Consider a differentiable function f(x) and its first derivative that we call p.

$$f\left(x\right) f'\left(x\right)=p\left(x\right)$$

The Legendre transformation that we call g(p) is the function whose first derivative is up to sign the inverse of the first derivative of the original function.

$$g\left(p\right) g'\left(p\right)=\pm x\left(p\right)$$

$$\Leftrightarrow g'\left(f\left(x\right)\right)=\pm x$$

We start by taking the total differential of g which happens to be plus minus xdp

$$dg=\frac{\partial g}{\partial p}dp=\pm xdp$$

And the total differentiable of F which happens to be pdx

$$df=\frac{\partial f}{\partial x}dx=pdx$$

Now we take the term p times x and compute its total differential and note that I put a plus minus sign in front just so that we'll get both variants of the the transformation with a single calculation.

$$\pm d\left(px\right)=\pm\left(\frac{\partial\left(px\right)}{\partial x}dx+\frac{\partial\left(px\right)}{\partial p}dp\right)$$

Evaluating these partial derivatives gives

$$\pm\left(pdx+xdp\right)=\pm df+dg$$

We can draw that d to the front since the total differential is linear and by comparison or integration if you will, we obtain that

$$\pm d\left(f\pm g\right)\Leftrightarrow px=f\pm g\Leftrightarrow g=\pm\left(px-f\right)$$

Application In Machine Learning

- Duality in Optimization

  In many ML problems, we have a primal problems (the org. problem) and a dual problem (the transformed problem).
  LF transform helps us to move between these two problems.

- Regularization

  Transform helps in formulating regularization terms

- Convex functions

  LF transform works well with convex functions, make it easier to analyze and solve them.