Duality (Boyd, Ch. 5)

Consider a constrained minimization problem

$$\begin{aligned} \min_x \ & f_0(x) \\ \text{s.t. } & f_i(x) \le 0 \quad , \quad i = 1, \dots, n \\ & h_j(x) = 0 \quad , \quad j = 1, \dots, m \end{aligned}$$

The Lagrangian is defined as

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) + \sum_{j=1}^{m} \nu_j h_j(x) .$$

λ_i, ν_j are called the Lagrange multipliers.

Define the Lagrange Dual function as

$$g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)$$

$\inf \mathcal{L}$ vs $\min \mathcal{L}$: Let $f(x) = \frac{1}{x}$. Then, $\forall x > 0$, $\inf_x \frac{1}{x} = 0$, but $\min_x \frac{1}{x}$ is undefined.

- if $\mathcal{L}$ is unbounded below on x, then $g(\lambda, \nu) = -\infty$.

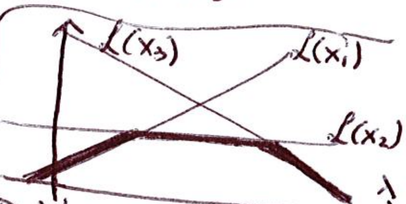
  For example: $\min_x c^T x$, with no constraint.

- $g(\lambda, \nu)$ is concave in (λ, ν):

  $\mathcal{L}$ is affine in (λ, ν), so it is concave (and convex).

  $\inf_x \mathcal{L}$ of a concave function is concave

  This happens even if $f_0(x)$ & $f_i(x)$ are not convex.



pointwise min of affine is concave.

So $$\boxed{g(\lambda,\nu) = \begin{cases} -b^T\nu, & \text{if } c + A^T\nu - \lambda = 0 \\ -\infty, & \text{else.} \end{cases}}$$

- $g(\lambda,\nu)$ is ~~a~~ affine in (λ,ν).
- the Lagrange dual of ~~the~~ a linear programming problem is a linear problem.

Lower bound on primal optimal value:

Claim Let $p^* = f_0(x^*)$, where x^* is the optimal solution of the primal problem.

Then $\boxed{g(\lambda,\nu) \leq p^*}$, for any $\lambda \geq 0$, any ν.

proof: Let $\tilde{x}$ be a feasible point. Then

$$f_i(\tilde{x}) \leq 0 \quad \text{and} \quad h_j(\tilde{x}) = 0.$$

So $$\sum_{i=1}^{n} \underbrace{\lambda_i}_{\geq 0} \underbrace{f_i(\tilde{x})}_{\leq 0} + \sum_{j=1}^{m} \nu_j \underbrace{h_j(\tilde{x})}_{=0} \leq 0$$

Hence

$$\mathcal{L}(\tilde{x},\lambda,\nu) = f_0(\tilde{x}) + \underbrace{\sum_{i=1}^{n} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{m} \nu_j h_j(\tilde{x})}_{\leq 0}$$

$$\leq f_0(\tilde{x})$$

Since $g(\lambda,\nu) \overset{\text{def}}{=} \inf_x \mathcal{L}(x,\lambda,\nu)$,

and $\inf_x \mathcal{L}(x,\lambda,\nu) \leq \mathcal{L}(\tilde{x},\lambda,\nu)$,

we have

$$g(\lambda,\nu) \leq \mathcal{L}(\tilde{x},\lambda,\nu) \leq f_0(\tilde{x}).$$

if ~~g(\lambda~~ $g(\lambda,\nu) > -\infty$, then says $\boxed{\text{dual feasible}}$

This holds for any feasible point $\tilde{x}$, so in particular holds for x^*. Therefore

$$g(\lambda,\nu) \leq f_0(x^*) = p^*.$$

<u>Example</u> Determine the Lagrange dual function of

$$\min_x \ \|x\|^2$$
$$\text{s.t.} \quad Ax = b$$

The Lagrange function is

$$\mathcal{L}(x, \nu) = \|x\|^2 + \nu^T(Ax - b).$$

In order to find the Lagrange dual function $g(\nu)$, we need to find

$$g(\nu) = \inf_x \mathcal{L}(x, \nu).$$

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + A^T\nu = 0 \Rightarrow \boxed{x = -\frac{1}{2}A^T\nu}$$

Therefore,

$$g(\nu) = \|x^*\|^2 + \nu^T(Ax^* - b)$$
$$= \left\|-\frac{1}{2}A^T\nu\right\|^2 + \nu^T\left(A\left(-\frac{1}{2}A^T\nu\right) - b\right)$$
$$= \frac{1}{4}\nu^T A A^T \nu - \frac{1}{2}\nu^T A A^T \nu - b^T\nu$$

$$\boxed{g(\nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T\nu.}$$

- $g(\nu)$ is concave in ν.

<u>Example</u>

$$\min \ c^T x$$
$$\text{s.t.} \quad Ax = b$$
$$x \geq 0$$

The Lagrange function is

$$\mathcal{L}(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T(Ax - b)$$
$$= -b^T\nu + (c + A^T\nu - \lambda)^T x$$

The Lagrange dual function is

$$g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)$$
$$= \cancel{\inf} \ -b^T\nu + \begin{cases} -\infty, & \text{if } c + A^T\nu - \lambda \neq 0 \\ 0, & \text{if } c + A^T\nu - \lambda = 0 \end{cases}$$

Lagrange Dual Problem

Since $g(\lambda,\nu) \le p^*$, we ask: What is the best upper bound?
This leads to the Dual Problem.

$$\max_{(\lambda,\nu)} \; g(\lambda,\nu) \quad \text{s.t.} \quad \lambda \ge 0.$$

Example:

(Primal) $\min_x \|x\|^2$ s.t. $Ax = b$

$g(\nu) = \frac{-1}{4}\nu^T A A^T \nu - b^T \nu$

$\boxed{x = \frac{-1}{2} A^T \nu}$

$\rightarrow$ (Dual) $\max_\nu g(\nu)$

$\Rightarrow \boxed{\max_\nu \frac{-1}{4}\nu^T A A^T \nu - b^T \nu}$

This is an unconstrained optimization.

Example:

(Primal) $\min c^T x$ s.t. $Ax = b$, $x \ge 0$ $\Rightarrow$ (Dual) $\max -b^T\nu$ s.t. $A^T\nu + c - \lambda = 0$, $\lambda \ge 0$

$$\Rightarrow \boxed{\begin{array}{l} \min b^T \nu \\ \text{s.t. } A^T\nu + c \ge 0 \end{array}}$$

- The dual problem is always convex, even if the primal is not convex: $g(\lambda,\nu)$ is concave. So $\max_{(\lambda,\nu)} g(\lambda,\nu)$ is a convex problem. $\lambda \ge 0$ is convex.
- If primal is also convex, then (possibly) the dual will solve the primal. (Need more conditions)

Define <u>duality gap</u>:

Let $d^* = \max_{\lambda,\nu} g(\lambda,\nu)$

s.t. $\lambda \geq 0$

Then since $g(\lambda,\nu) \leq p^*$ $\forall \lambda \geq 0$ and ν, we also have

$$\boxed{d^* \leq p^*}$$

The gap is called the $\boxed{\text{duality gap} = p^* - d^*}$.

<u>Strong Duality</u>

If the primal problem is $\left\{ \begin{array}{l} \min f_0(x) \\ \text{s.t. } f_i(x) \leq 0 \\ \quad Ax = b \end{array} \right.$ i.e. $h(x)$ is affine,

then when <u>Slater's constraint qualification</u> is satisfied, i.e.

$\left\{ \begin{array}{l} \text{there exists } x \text{ s.t.} \\ \quad f_i(x) < 0, \quad i = 1, 2, \ldots, n, \\ \text{and } Ax = b \end{array} \right.$

This says there are points that are strictly feasible!

then

$d^* = p^*$ if primal is convex.

<u>Message</u>: When primal is convex, then we can solve the dual and get back the primal solution.

When primal is not convex, then the dual solution would be a convex approximation to the primal.

Dual Norm and Conjugate function

Let $\|x\|$ be the norm of x. The dual norm of $\|\cdot\|$ is

$$\|x\|_* = \max_{\|z\|\le 1} (z^T x).$$

Properties: (1) For $\|x\|_2$, the dual norm is $\|x\|_* = \|x\|_2$

(By Cauchy inequality) $(z^T x)^2 \le \|z\|^2 \|x\|^2$

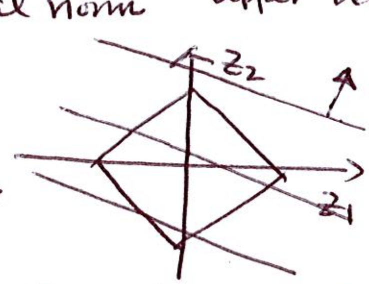
$\le \|x\|^2 \; \forall z.$

Choose $z = \frac{x}{\|x\|}$, then upper becomes tight.

(2) For $\|x\|_1$, the dual norm is $\|x\|_* = \|x\|_\infty$

$$x^T z = \sum_i x_i z_i \le \sum_i |x_i||z_i| \le \max_i |x_i| \left(\sum_i |z_i|\right) = \max_i |x_i|$$

$$\max_z \; z^T x \quad \text{s.t. } \|z\|_1 \le 1$$



slope defined by x.

When max is attained, only one of the vertices is activated. This corresponds to $\max(x_1, x_2)$.

(3) $\|x\|_{**} = \|x\|$. So the dual norm of $\|x\|_\infty$ is $\|x\|_1$.

Let $f: \mathbb{R}^n \to \mathbb{R}$. The function $f^*: \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) = \sup_x \{ y^T x - f(x) \},$$

is called the conjugate function.

Example (i) if $f(x) = \|x\|_1$, then (See Boyd Example 3.26)

$$f^*(y) = \begin{cases} 0, & \text{if } \|y\|_\infty \le 1 \\ \infty, & \text{else.} \end{cases}$$

(ii) if $f(x) = \log\left(\sum_{i=1}^n e^{x_i}\right)$, then (See Boyd Example 3.25)

$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i, & \text{if } y \ge 0,\ \mathbf{1}^T y = 1 \\ \infty, & \text{else.} \end{cases}$$

Solving Primal Problems Via Dual

Consider $\min_x f_0(Ax+b)$

The Lagrange dual is

$$g(\lambda, \nu) = \inf_x \{ f_0(Ax+b) + \lambda^T 0 + \nu^T 0 \}$$
$$= p^* .$$

So even though we have $p^* = d^*$ (i.e. strong duality), this dual problem is not informative.

Alternatively, consider

$$\min_{x,y} f_0(y)$$
$$\text{s.t. } Ax + b = y$$

$$L(x,y,\nu) = f_0(y) + \nu^T(Ax+b-y)$$

$$\frac{\partial L}{\partial x} = 0 \Rightarrow A^T\nu = 0$$

The Lagrange dual is

$$g(\nu) = \inf_{x,y} L(x,y,\nu) = \inf_y \{ \inf_x L(x,y,\nu) \}$$

$$= \inf_y \begin{cases} f_0(y) + \nu^T b - \nu^T y, & A^T\nu = 0 \\ -\infty & , A^T\nu \neq 0 \end{cases}$$

$$= \begin{cases} \inf_y \{ f_0(y) - \nu^T y \} + b^T\nu, & A^T\nu = 0 \\ -\infty & , A^T\nu \neq 0 \end{cases}$$

$$= \begin{cases} b^T\nu - f_0^*(\nu) & , A^T\nu = 0 \\ -\infty & , A^T\nu \neq 0 . \end{cases}$$

So the dual problem is

$$\max_\nu \ b^T\nu - f_0^*(\nu)$$
$$\text{s.t. } A^T\nu = 0$$

Example

$$\min_x \log\left(\sum_{i=1}^m \exp\{a_i^T x + b\} \right)$$

$$f_0(y) = \log\left(\sum_{i=1}^m \exp y_i \right)$$

$$f_0^*(\nu) = \begin{cases} \sum_{i=1}^m \nu_i \log \nu_i, & \nu \geq 0, \ \nu^T 1 = 1 \\ \infty & , \text{else} \end{cases}$$

$$\Rightarrow \min_{x,y} \log\left(\sum_{i=1}^m \exp y_i \right)$$
$$\text{s.t. } Ax + b = y$$

$\Rightarrow$ dual problem:

$$\boxed{\begin{aligned} &\max_\nu -\sum_{i=1}^m \nu_i \log \nu_i + b^T\nu \\ &\text{s.t. } 1^T\nu = 1, \quad \nu \geq 0, \quad A^T\nu = 0 . \end{aligned}}$$

~~(If A = I and b = 0, then the dual has closed form)~~

Pictorial Illustration of Conjugate Function

Consider the conjugate function:

$$f^*(y) = \sup_x \{ yx - f(x) \}.$$

find x s.t. $yx - f(x)$ is maximized

$\Leftrightarrow$ find x s.t. $\nabla f(x) = y$.

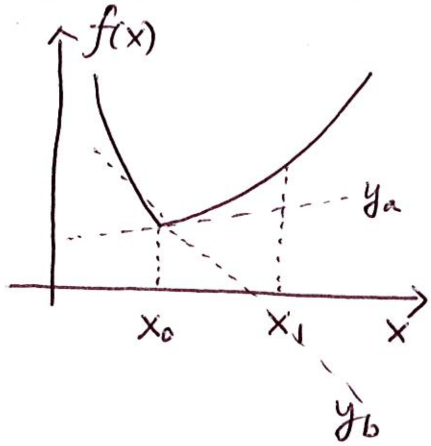
Three properties of $f^*(y)$:

(i) $f^*(y)$ is always convex

(ii) $f^{**}(x)$ is always convex

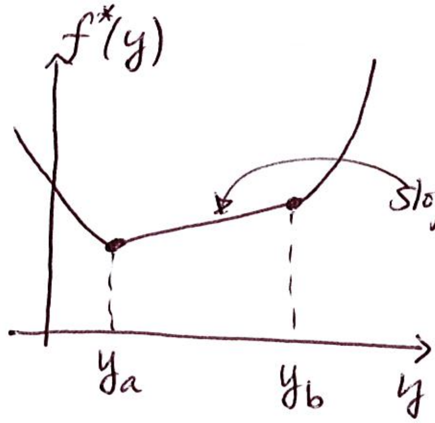
(iii) $f^{**}(x) = f(x)$ if $f(x)$ is convex.

Example f convex but not differentiable



At x_1, there is only one slope that ~~make $\nabla f(x) = y$, which is~~ can maximize $yx - f(x)$, this slope y is $y = f'(x_1)$.

At x_0, there are multiple slopes that can maximize $yx - f(x)$.
Let's say this range of slope is $[y_a, y_b]$.
Then there will be a line in ~~the~~ $f^*(y)$



<u>Proofs</u> (i)

$$f^*(\alpha y_1 + (1-\alpha)y_2)$$
$$= \sup_x \{ (\alpha y_1 + (1-\alpha)y_2)x - f(x) \}$$
$$= \sup_x \{ \alpha y_1 x - \alpha f(x) \cancel{\}} + (1-\alpha)y_2 x - (1-\alpha)f(x) \}$$
$$\leq \sup_x \{ \alpha (y_1 x - f(x)) \} + \sup_x \{ (1-\alpha)(y_2 x - f(x)) \}$$
$$= \alpha f^*(y_1) + (1-\alpha) f^*(y_2).$$

(ii) Follows from (i).

(iii) Define $f^{**}(x) = \sup_z \{ xz - f^*(z) \}$.

Let $\hat{z}$ be the maximizer :

$$\hat{z} = \underset{z}{\operatorname{argmax}} \{ xz - f^*(z) \}.$$

Then $f^{**}(x) = x\hat{z} - f^*(\hat{z})$.

Since $\hat{z}$ is the maximizer, we have

$$\frac{d}{dz}\left(xz - f^*(z) \right)\Big|_{z=\hat{z}} = 0 \implies x = f^{*\prime}(\hat{z}).$$

So $f^{**}(x) = f^{*\prime}(\hat{z})\hat{z} - f^*(\hat{z})$

$$= f^{*\prime}(\hat{z})\,\hat{z} - \sup_u \{ u\hat{z} - f(u) \}$$

Let $\hat{u} = \underset{u}{\operatorname{argmax}} \{ u\hat{z} - f(u) \}$. Then

$$\sup_u \{ u\hat{z} - f(u) \} = \hat{u}\hat{z} - f(\hat{u})$$

So $f^{**}(x) = f^{*\prime}(\hat{z})\hat{z} - \hat{u}\hat{z} + f(\hat{u})$.

Now, if we can show that

~~(a) $\hat{u} \le \hat{z}$~~

~~(b)~~ $\hat{u} = f^{*\prime}(\hat{z})$,

then we are done.

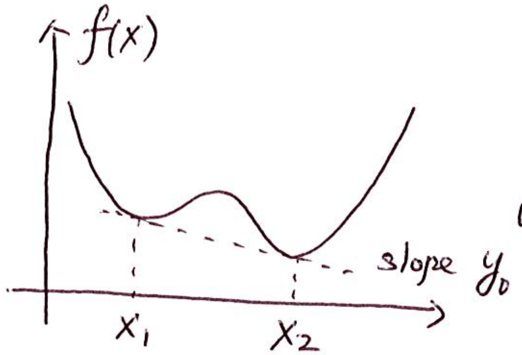
To this end, note that

$$f^{*\prime}(z) = \frac{d}{dz} f^*(z) = \frac{d}{dz}\left(\max_u \{uz - f(u)\}\right)$$

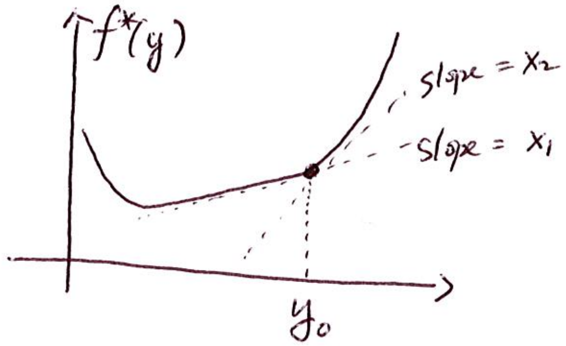
$$= \frac{d}{dz}\left(\hat{u} z - f(\hat{u})\right) = \hat{u}.$$

~~And since $\hat{z} \le$~~

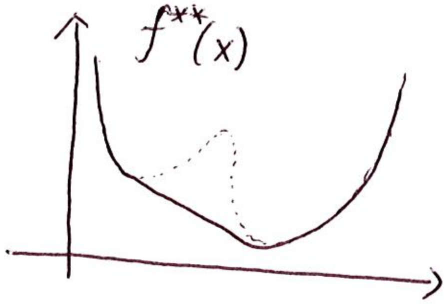
Example f non-convex



if f is ~~con~~ non-convex, there ~~∄~~ is a line connecting x_1 and x_2. let slope be y_0.



On $f^*(y)$, at y_0, there are two slopes x_1 and x_2. ~~In fact for any slope x ∈ [x1, x2].~~



$f^{**}(x)$ is the convex envelope of $f(x)$ if f is non-convex.