Duality Gaps, complementary Slackness & KKT conditions:

Recall: Want to solve

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g_i(x) \le 0 \qquad i = 1, \dots, m \\ &\quad\ h_j(x) = 0 \qquad j = 1, \dots, p \end{aligned}$$

— Primal opt. prob.

So we write the Lagrangian

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

The Lagrangian Dual function is

then $$F(\lambda, \nu) = \min_{x \in D} L(x, \lambda, \nu)$$

& the dual opt. problem is

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

Recall: Weak duality

$\Rightarrow \quad \beta^* \leq \alpha^*$ always holds (i.e. for all opt. problems)

β^*: max of dual problem

α^*: min of primal problem

Strong duality

$$\alpha^* = \beta^*$$

Does not always hold, but it sometimes does (e.g. when Slater's conditions hold)

Duality Gap

Sometimes solving the dual problem is easier than solving the primal

Even more, since $\beta^* \leq \alpha^*$

β^*: max of (D)

α^*: min of (P)

Then for any dual-feasible (λ, ν) & primal feasible x:

$$F(\lambda,\nu) \leq \beta^* \leq \alpha^* \leq f(x)$$

$$\Rightarrow \boxed{f(x) - \alpha^* \leq f(x) - F(\lambda,\nu)}$$

so if $f(x) - F(\lambda,\nu)$ is small then we know that $f(x) - \alpha^*$ is small (i.e. x is close to being optimal)

Defn: For an opt. problem with objective f & dual F, primal feasible x & dual feasible (λ,ν) the quantity

$$f(x) - F(\lambda,\nu)$$

is known as the duality gap

Usefulness: Suppose you have an algorithm that produces a series of primal & dual feasible points

$$x^{(i)} \quad \& \quad (\lambda^{(i)}, \nu^{(i)})$$

then if $f(x^{(i)}) - F(\lambda^{(i)}, \nu^{(i)}) < \varepsilon$

we also have $f(x^{(i)}) - d^* < \varepsilon$

Complementary Slackness:

Lemma: Let $x^* \in \mathbb{R}^n$ be primal optimal & let (λ^*, ν^*) be dual optimal & suppose that we have strong duality, then

(1) x^* minimizes $L(x, \lambda^*, \nu^*)$

(2) $\lambda_i^* g_i(x^*) = 0 \quad \forall i = 1, \ldots, m$

↳ complementary slackness

$\lambda_i^* \; g_i(x^*) = 0 \quad \forall i$

$\Rightarrow$ when $\lambda_i^* > 0$, $g_i(x^*) = 0$

when $g_i(x^*) < 0$, $\lambda_i^* = 0$

Proof of Lemma:

(1) $f(x^*) = F(\lambda^*, \nu^*)$ by strong D.

$= \min_x f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) + \sum_{j=1}^p \nu_j^* h_j(x)$ (by def'n of F)

$\leq f(x^*) + \sum_i \lambda_i^* g_i(x^*) + \sum_j \nu_j^* h_j(x^*)$ (pick $x = x^*$)

(≥ 0, ≤ 0, 0)

$\leq f(x^*)$

But (1) $f(x^*) = f(x^*)$ (of course)

so the inequalities should have all been equalities, i.e.

$$\min_x \underbrace{f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) + \sum_{j=1}^p \nu_j^* h_j(x)}_{L(x, \lambda^*, \nu^*)} = f(x^*)$$

(which is what we needed in order to prove (1))

and $f(x^*) + \sum_i \lambda_i^* g_i(x^*) + \sum_j \nu_j^* h_j(x^*) = f(x^*)$

$\lambda_i^* \geq 0$, $g_i(x^*) \leq 0$, $\sum_i \lambda_i^* g_i(x^*) \leq 0$; $\sum_j \nu_j^* h_j(x^*) = 0$

so this has to be zero which proves (2)

Karush-Kuhn-Tucker (KKT) conditions

These are necessary & sufficient conditions for strong duality

Defn: $x^* \in \mathbb{R}^n$ & $(\lambda^*, \nu^*) \in \mathbb{R}^m \times \mathbb{R}^p$ satisfy the KKT cond's if

1. $g_i(x^*) \leq 0 \quad \forall i = 1, \ldots, m$
2. $h_i(x^*) = 0 \quad \forall i = 1, \ldots, p$
3. $\lambda_i \geq 0 \quad \forall i = 1, \ldots, m$
4. $\lambda_i^* g_i(x^*) = 0 \quad \forall i = 1, \ldots, m$

and

5. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$

Theorem:

Nec. Cond.

(1) Strong duality $\Rightarrow$

primal optimal pt. $x^* \in \mathbb{R}^n$ &
dual optimal pts $(\lambda^*, \nu^*) \in \mathbb{R}^m \times \mathbb{R}^p$

satisfy the KKT conditions

Suff. Cond.

(2) If f & g_i are convex & if h_i are all affine and we have the KKT conditions $\Rightarrow$ strong duality holds

Proof: (1) The nec. cond:

- We have ①, ②, ③ by feasibility of x^* & (λ^*, ν^*)
- We have ④ by the Lemma
- We have ⑤ since x^* minimizes $L(x, \lambda^*, \nu^*)$
$\Rightarrow \nabla_x L(x, \lambda^*, \nu^*)\big|_{x^*} = 0$.

(2). KKT cond's ① & ② $\Rightarrow x^*$ is primal feasible

- KKT cond ③ $\Rightarrow (\lambda^*, \nu^*)$ is dual feasible
- f, g_i convex & h_i affine

$\Rightarrow L(x, \lambda^*, \nu^*) = f(x) + \sum \lambda_i^* g_i(x) + \sum \nu_j h_j(x)$ is convex in x.

- KKT cond ⑤ $\Rightarrow x^*$ is the global minimizer of $L(x, \lambda^*, \nu^*)$

$\Rightarrow F(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$

$= f(x^*) + \underbrace{\sum \lambda_i^* g_i(x^*)}_{0 \text{ by } ④} + \underbrace{\sum \nu_j h_j(x^*)}_{0 \text{ by } ②}$

$\Rightarrow F(\lambda^*, \nu^*) = f(x^*)$

But $F(\lambda^*, \nu^*) \leq f(x^*)$ by weak D so (λ^*, ν^*) maximizes F & we have strong duality ∎

- We already saw one algorithm for solving constrained opt. problems, namely Projected gradient descent.
  (needed to know how to compute projections).

- Another way for solving non-const. opt. problems is to try to (numerically) solve the KKT conditions which is a system of equations & inequalities with

  $n+m+p$ unknowns.

- Idea: $\min_{x \in \mathbb{R}^n} f(x)$

  s.t. $g_i(x) \leq 0 \quad i = 1, \dots, m$

  $\Rightarrow L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$

  "Lagrangian algorithm":

$$\begin{cases} x^{(t+1)} = x^{(t)} - \alpha_t \left(\nabla f(x^{(t)}) + \sum_{j=1}^{m} \lambda_i \nabla g_i(x^{(t)}) \right) \\ \lambda^{(t+1)} = \Pi_+ \left(\lambda^{(t)} + \beta_t g(x^{(t)}) \right) \end{cases}$$

where $\Pi_+(\vec{a}) = \vec{b} \in \mathbb{R}^m$

with $\begin{cases} b_i = a_i & \text{if } a_i \geq 0 \\ b_i = 0 & \text{if } a_i < 0 \end{cases}$

→ Projection onto +ve orthant
i.e. the set $\{\lambda_i \geq 0, i = 1, \ldots, m\}$

Above, the x update is a GD step
the λ update is a projected gradient ascent step.

Intuition: want to minimize $f(x)$
& maximize $F(\lambda)$.

Examples on duality / KKT:

minimize $x_1^2 + x_2^2$

s.t. $\begin{cases} x_1 + x_2 \geq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$ $\rightarrow D = \{x$ s.t. $x_1 + x_2 \geq 4,\ x_1 \geq 0, x_2 \geq 0\}$

Then $\mathcal{L}(x, \lambda) = x_1^2 + x_2^2 + \lambda_1(4 - x_1 - x_2) + \lambda_2(-x_1) + \lambda_3(-x_2)$

$$F(\lambda) = \min_{x \in D} \; x_1^2 + x_2^2 + \lambda_1(4 - x_1 - x_2) + \lambda_2(-x_1) + \lambda_3(-x_2)$$

$$= 4\lambda_1 + \min_{x \in D} \underbrace{(x_1^2 - \lambda_1 x_1 - \lambda_2 x_1)} + (x_2^2 - \lambda_1 x_2 - \lambda_3 x_2)$$

minimizing $\rightarrow \hat{x}_1 = \frac{\lambda_1 + \lambda_2}{2}$, $\hat{x}_2 = \frac{\lambda_1 + \lambda_3}{2}$ (why?)

$$\Rightarrow F(\lambda) = 4\lambda_1 - \frac{(\lambda_1 + \lambda_2)^2}{4} - \frac{(\lambda_1 + \lambda_3)^2}{4}$$

$$\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0$$

$F(\lambda)$ is concave & differentiable and we want to maximize it

Since $\lambda_1, \lambda_2 \geq 0$, $\lambda_3 \geq 0$, by inspection we see that to maximize $F(\lambda)$ $\lambda_2^* = 0$ & $\lambda_3^* = 0$ (complementary slackness!!)

$\Rightarrow$ $F(\lambda)$ reduces to

$$F(\lambda_1) = 4\lambda_1 - \frac{\lambda_1^2}{2} \Rightarrow \lambda_1^* = 4$$

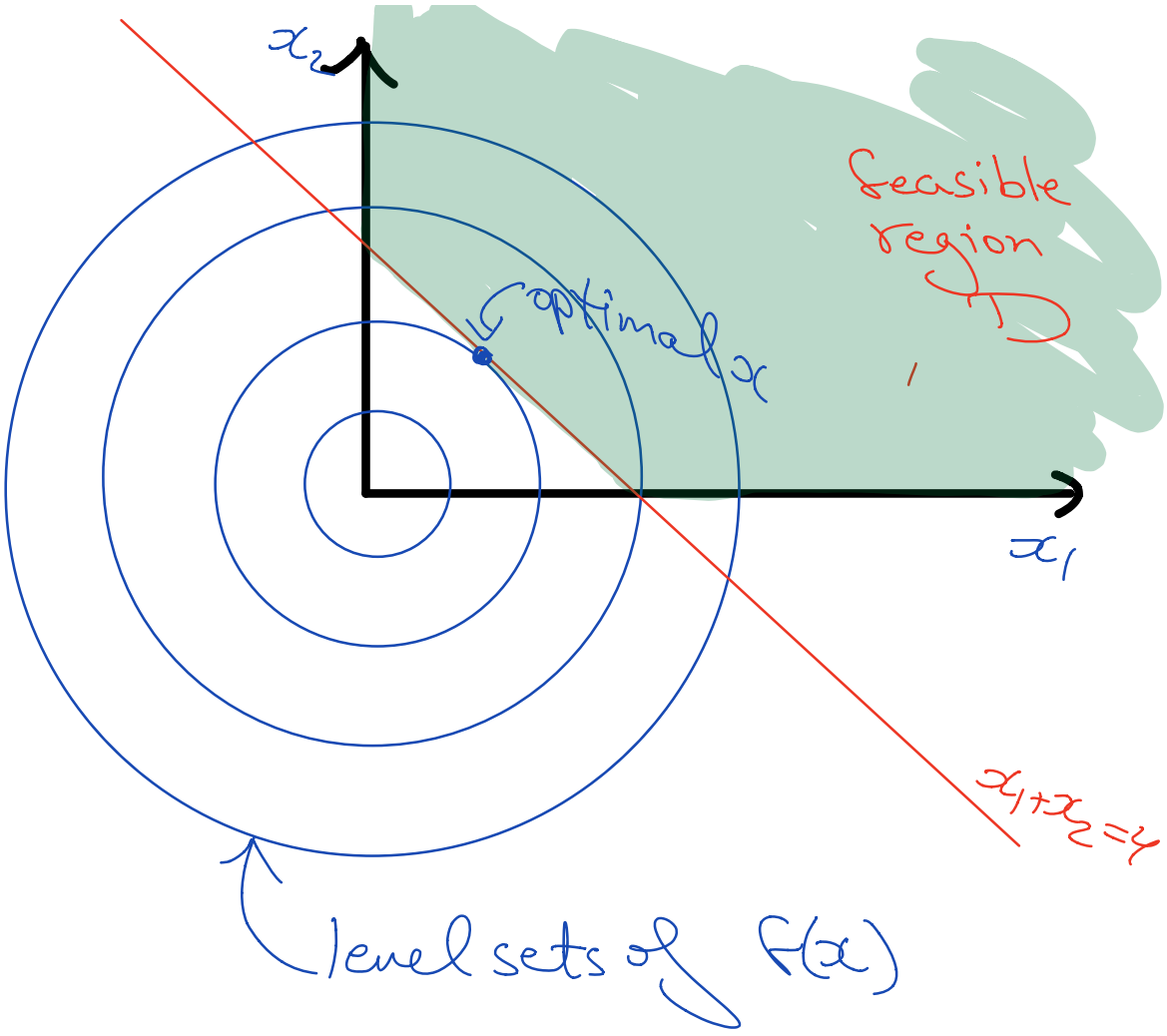
$$\Rightarrow \boxed{F(\lambda^*) = 8}$$

on the other hand

$$\hat{x}_1 = \frac{\lambda_1^* + \lambda_2^*}{2} = 2$$

$$\hat{x}_2 = 2$$

$\Rightarrow$ $(2, 2)$ is opt.



Notice that at the optimal soln

$$x_1^* + x_2^* - 4 = 0, \quad \lambda_1^* \neq 0$$

$$x_1^* > 0, \quad \lambda_2^* = 0$$

$$x_2^* > 0, \quad \lambda_3^* = 0$$

(Complementary slackness)

Exercise: minimize x_1

$$\text{s.t. } x_1^2 + x_2^2 = 1$$

- What is the soln to this opt. problem?
- Formulate the dual.
- Solve it
- Is $\alpha^* = \beta^*$?