Lecture 22 Optimality Conditions

Note by Samion Suwito on 4/10/25

Last Time

Lagrange Duality

A descriptive framework for forming dual problem of constrained optimisation problem
Primal Problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f_0(x) \\ &\text{s.t. } f_i(x) \le 0 \quad i = 1\ldots m \end{aligned} \tag{P}$$

$\implies$ *Dual Problem*

$$\begin{aligned} \max_{\lambda\ge 0}(\inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda)) \\ \mathcal{L}(x,\lambda) = f_0(x) + \sum_i \lambda_i f_i(x) \end{aligned} \tag{D}$$

$p^\star \ge d^\star$, optimal dual variable = "**Lagrange Multiplier**"

Strong Duality holds if, e.g. there exists strictly feasible primal x meaning $(f_i(x) \le 0 \ \ i = 1\ldots m)$ And all f_i's are convex. (**Slater's Condition**)

Dual Problems for Standard Problem Classes

Standard Problem Classes (LP, QP, etc.) have established dual problems for primal problems in standard form..

Ex. LP Duality

$A \in \mathbb{R}^{m\times n}$
Primal Problem:

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} c^\top x \\ &\text{s.t. } Ax \le b \end{aligned} \tag{P}$$

Dual Problem:

$$\begin{aligned} \max_{y \in \mathbb{R}^m} \; & -b^\top y \\ \text{s.t. } & A^\top y = -c \\ & y \geq 0 \end{aligned} \tag{D}$$

Using Lagrange duality:
$\mathcal{L}(x, \lambda) = c^\top x + \sum \lambda_i (a_i^\top x - b_i)$

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) &= c + \sum \lambda_i a_i \\ &= \begin{cases} 0 \text{ only when } A^\top \lambda = -c \\ \text{nonzero otherwise} \end{cases} \end{aligned}$$

We can then use this to show

$$\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = \begin{cases} -\sum \lambda_i b_i \text{ if } A^\top \lambda = -c \\ -\infty \text{ otherwise} \end{cases}$$

$$\begin{aligned} \max_{\lambda \geq 0} (\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda)) = \max_{\lambda} & \left(-\sum \lambda_i b_i \right) \\ \text{s.t. } & \lambda \geq 0 \\ & A^\top \lambda = -c \end{aligned}$$

Which if we relabel λ to y gives us the dual problem

Theorem: (Strong duality for LPs) If either (*P*) or (*D*) are feasible, then strong duality holds.

(CS170) Max flow is actually dual to Min cut which is why they have The same optimal value

Strong Duality not holding

$$\begin{aligned} \min_{x, y \in \mathbb{R}} \; & e^{-x} + I_{(0,\infty)}(y) \\ \text{s.t. } & \frac{x^2}{y} \leq 0 \end{aligned}$$

The constraint is convex on $\{(x, y) \in \mathbb{R} \times (0, \infty)\}$, We can clearly see the optimal value by just setting x to 0 and y to as close to 0 making $p^\star = 1 = e^{-0}$. Create the dual problem using lagrange duality:

$$\inf_{x,y} \left\{ e^{-x} + I_{(0,\infty)}(y) + \lambda \frac{x^2}{y} \right\} = \begin{cases} 0 \text{ if } \lambda \geq 0 \\ -\infty \text{ if } \lambda < 0 \end{cases} \implies d^\star = 0$$

However rewriting to

$$\begin{aligned}&\min_x e^{-x}\\&\text{s.t. } x = 0\end{aligned}$$

Does satisfy strong duality

Problems with Equality Constraints

Consider primal problem:

$$\begin{aligned}&\min_x f_0(x)\\&\text{s.t. } f_i(x) \leq 0 \quad i = 1\dots m\\&\qquad h_j(x) = 0 \quad j = 1\dots k\end{aligned} \tag{P}$$

$\implies$

$$\begin{aligned}&\min_x f_0(x)\\&\text{s.t. } f_i(x) \leq 0 \quad i = 1\dots m\\&\qquad h_j(x) \leq 0 \quad j = 1\dots k\\&\qquad - h_j(x) \geq 0 \quad j = 1\dots k\end{aligned}$$

Then

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{k} (\lambda_{m+j} - \lambda_{m+k+j}) h_j(x)$$

Giving the dual problem:

$$\max_{\lambda \geq 0} \inf_x \mathcal{L}(x, \lambda) = \max_{\lambda \geq 0, \mu \in \mathbb{R}^k} \inf_x \mathcal{L}(x, \lambda, \mu)$$

where

$$\mathcal{L}(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

Remark: If the primal problem is a convex optimisation, then the equality constraint h_j functions must be affine (as h_j and $-h_j$ must be convex). So, $h_j(x) = 0, j = 1\dots k$ can be rewritten as $Ax = b$ for some A and b.

Certificate of Optimality

Define: $g(\lambda, \mu) = \inf_{\lambda \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$
If x is primal feasible, and (λ, μ) dual-feasible, then $f_0(x) \geq p^\star \geq d^\star \geq g(\lambda, \mu)$ by definition. We can then conclude

$$f_0(x) - p^\star \leq \underbrace{f_0(x) - g(\lambda, \mu)}_{\text{duality gap}}$$

If the gap is small then x is near optimal. Since duality gap can approach zero when strong duality holds, strong duality gives stopping criteria. This is useful as we can set a threshold to when to end a program.

KKT Conditions

KKT Conditions are “first order” necessary conditions for optimality.

If all f_i’s and h_j’s are differentiable (no convexity assumptions), then $x^\star$ primal optimal and $(\lambda^\star, \mu^\star)$ dual optimal and $p^\star = d^\star$ must satisfy:

1. $\nabla f_0(x^\star) + \sum \lambda_i^\star \nabla f_i(x^\star) = \sum \mu_i^\star \nabla h_j(x^\star) = 0$ (stationarily?)
2. $x^\star$ is primal feasible; $(\lambda^\star, \mu^\star)$ is dual feasible
3. $\lambda^\star f_i(x^\star) = 0 \quad i = 1 \ldots m$ (complementary slackness)

Proof: Claim since $p^\star = d^\star$, $x^\star$ minimises $x \mapsto \mathcal{L}(x, \lambda^\star, \mu^\star)$.

$$\begin{aligned} d^\star &= \inf_{x \in X} \left\{ f_0(x) + \sum \lambda_i^\star f_i(x) + \sum \mu_j^\star h_j(x) \right\} \\ &\leq f_0(x^\star) + \sum \lambda_i^\star \underbrace{f_i(x^\star)}_{\leq 0} + \sum \mu^\star \underbrace{h_j(x^\star)}_{=0} \\ &\leq f_0(x^\star) = p^\star = d^\star \end{aligned}$$

Using the first condition we can show how $x^\star$ minimises the first line showing equality for the second. Then using the second feasibility condition we can show the $f_i(x) \leq 0$ and $h_j(x) = 0$. Finally using the third condition we make it $\lambda_i^\star f_i(x^\star) = 0$ showing the last equality.

If strong duality holds, KKT conditions necessary for $x^\star$, $(\lambda^\star, \mu^\star)$ to be primal/dual optimal respectively.

Theorem: If f_i's convex differentiable, h_j's affine then KKT conditions also sufficient.

Proof: Suppose $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$ satisfies KKT conditions, by convexity and KKT Condition 1 and 2 $\tilde{x}$ minimises $x \mapsto \mathcal{L}(x, \tilde{\lambda}, \tilde{\mu})$. We know

$d^\star \geq \inf_x \mathcal{L}(x, \tilde{\lambda}, \tilde{\mu}) \underbrace{=}_{\text{KKT 3}} f_0(\tilde{x}) \underbrace{\geq}_{\text{KKT 2}} p^\star \underbrace{\geq}_{\text{weak duality}} d^\star$ meaning we get equality throughout.