Lecture 23 Examples of KKT Conditions

Note by Samion Suwito on 4/15/25

Last Time

Primal Dual

Primal Problem:

$$\begin{aligned} \min_{x\in\mathbb{R}^n}\ & f_0(x) \\ \text{s.t. } & f_i(x) \leq 0, \quad i = 1\dots m \\ & h_j(x) = 0, \quad j = 1\dots k \end{aligned} \tag{P}$$

Dual Problem:$g(\lambda, \mu) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$

$$\begin{aligned} \min_{\lambda\in\mathbb{R}^m, \mu\in\mathbb{R}^k}\ & g(\lambda, \mu) \\ \text{s.t. } & \lambda \geq 0 \end{aligned} \tag{D}$$

where $\mathcal{L}(x, \lambda, \mu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \mu_j h_j(x)$

KKT Conditions

For (x, λ, μ)

1. Feasibility:
   - $f_i(x) \leq 0, i = 1\dots m,$
   - $h_j(x) = 0, j = 1\dots k$
   - $\lambda_i \geq 0, i = 1\dots m$
2. Stationarity:
   - $\nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) + \sum_j \mu_j \nabla h_j(x) = 0$
3. Complementary Slackness: $\lambda_i f_i(x) = 0, i = 1\dots m$

Theorem:
1. If $x^\star$ optimal in (P), $(\lambda^\star, \mu^\star)$ optimal in (D), and strong duality holds $(p^\star = d^\star)$, then $(x^\star, \lambda^\star, \mu^\star)$ satisfy KKT conditions. (Essentially means KKT conditions are necessary.)
2. If $(f_i)_{i=0}^m$ are convex and $(h_j)_{j=1}^k$ are affine (i.e. (P) is a convex problem) and

$(x^\star, \lambda^\star, \mu^\star)$ satisfy KKT conditions, then $x^\star$ is optimal in (P), $(\lambda^\star, \mu^\star)$ is optimal in (D) and strong duality holds (i.e. $p^\star$=$d^\star$).

Examples of KKT

Example 1

$$\begin{aligned}&\min_x \frac{1}{2}x^\top Hx + c^\top x \quad H \succeq 0\\ &\text{s.t. } Ax = b\end{aligned}$$

$\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Hx + c^\top x + \mu^\top(Ax - b)$

KKT Conditions

1. Feasibility $Ax = b$
2. Stationarity: $Hx + c + A^\top\mu = 0$
3. No λ so $\lambda_i f_i(x = 0$

Based on KKT Conditions we can create the linear equation

$$\begin{bmatrix} H & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \mu \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Any solution to this equation is primal/dual optimal. If there is no solution than the primal optimal is not attained.

Example 2

Consider investment problem where we can invest money in n different assets. Covariance of assets given by $n \times n$ matrix $C \succeq 0$.

A portfolio is a vector $x \in \mathbb{R}^n$ where $x_i \geq 0$ is amount we invest in asset i.

$$\text{Portfolio Risk} = R = x^\top Cx = \sum_{i=1}^{n} \underbrace{x_i(Cx)_i}_{\text{risk of asset i}}$$

Portfolio x has risk-parity if $x_i(Cx)_i = \frac{1}{n}R \quad \forall i = 1 \ldots n$ essentially diversification. Consider:

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} \frac{R}{n} \sum_{i=1}^{n} \phi(x_i) + \frac{1}{2} x^\top C x \\
&\text{s.t. } x \geq 0 \quad i = 1 \dots n
\end{aligned}
$$

Where $\phi(t)$ is a log barrier function.

$$
\phi(t) = \begin{cases} \log\left(\frac{1}{t}\right) & t > 0 \\ +\infty & t \leq 0 \end{cases}
$$

We can solve by using the KKT conditions

$$
\mathcal{L}(x, \lambda) = \begin{cases} \frac{R}{n} \sum -\log(x_i) + \frac{1}{2} x^\top C x - \lambda^\top x & x > 0 \\ +\infty \text{ otherwise} \end{cases}
$$

KKT:

1. $x \geq 0, \lambda_i \geq 0 \quad \forall i = 1 \dots n$
2. $-\frac{R}{n} \frac{1}{x_i} + (Cx_i) - \lambda_i = 0, i = 1 \dots n$
3. $\lambda_i x_i = 0$

By KKT 1 and **Slater**, $(x^\star, \lambda^\star)$ satisfy KKT, must have $x_i^\star > 0$ (due to log barrier function) $\implies \lambda_i^\star = 0$ using complementary slackness. $\implies x_i^\star (Cx^\star)_i = \frac{R}{n}$ using KKT 2 therefore having risk parity.

Example 3

For a probability vector p. The entropy $H(p) = \sum_i p_i \log \frac{1}{p_i}$, which is a convex function of p.
(big entropy = randomness, small entropy = deterministic)

$$
\begin{aligned}
&\min_p \sum p_i \log p_i \\
&\text{s.t. } . p_i \geq 0 \\
&\sum p_i = 1 \\
a^\top p = &\sum a_i p_i \leq 0
\end{aligned}
$$

The lagrangian:

$$\mathcal{L}(p,\lambda,\nu,\mu) = \sum p_i \log p_i - \lambda^\top p + \mu\left(\sum p_i - 1\right) + \nu a^\top p$$

The KKT:

1. $\sum p_i = 1, p_i \geq 0, \lambda \geq 0, \nu \geq 0$
2. $\log p_i + 1 - \lambda_i + \mu + \nu a_i = 0$
3. $\lambda_i p_i = 0, \nu \sum a_i p_i = 0$

Using KKT2: $p_i = e^{-(\mu+1)-\nu a_i+\lambda_i} > 0$
$\implies \lambda_i = 0$ by KKT3
$\implies p_i = e^{-(\mu+1)-\nu a_i}$
$\implies 1 = \sum p_i = e^{-(\mu+1)} \sum_i e^{-\nu a_i}$ by KKT1, where μ serves to normalise:
$e^{\mu+1} = \sum_i e^{-\nu a_i}$ which makes the equation above 1.

From here he corrects his work look below to what he wrote before for more understanding:
$0 \geq \sum a_i p_i \implies \sum a_i e^{-\nu a_i} \leq 0$ through primal feasibility and $\nu \sum a_i e^{-\nu a_i} = 0$ by complementary slackness

> Pre-correction
>
> $\implies 0 = \nu \sum a_i p_i = \nu \sum a_i e^{-(\mu+1)} e^{-\nu a_i}$
> $\implies \nu \sum_i a_i e^{-\nu a_i} = 0$ where $\nu \geq 0$
> We got rid from $e^{-(\mu+1)}$ since it's positive and we divide by it.
> We also can't just set $\nu = 0$ and must satisfy $\sum a_i e^{-\nu a_i} \leq 0$ primal feasibility