FOUNDATIONS OF OPTIMIZATION: IE6001

Optimality Conditions

Napat Rujeerapaiboon

Semester I, AY2022/2023

Complementary Slackness

Assume that strong duality holds and that $x^\star$ is optimal in P, while $(\lambda^\star, \mu^\star)$ is optimal in D. Then,

$$\begin{aligned}
f_0(x^\star) &= g(\lambda^\star, \mu^\star) && \text{(strong duality)} \\
&= \inf_x f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \mu_i^\star h_i(x) && \text{(definition of } g\text{)} \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \mu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star) && (\lambda^\star \geq 0,\ f(x^\star) \leq 0,\ h(x^\star) = 0)
\end{aligned}$$

and thus all inequalities hold as equalities. We conclude:

- $x^\star$ minimizes $L(x, \lambda^\star, \mu^\star)$
- $\sum_{i=1}^m \lambda_i^\star f_i(x^\star) = 0$, which implies *complementary slackness*, i.e.,

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0 \quad \text{and} \quad f_i(x^\star) < 0 \implies \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) Conditions

Theorem: Assume that $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable, $x^\star$ is optimal in P, $(\lambda^\star, \mu^\star)$ is optimal in D and strong duality holds (but P may be nonconvex). Then, $(x^\star, \lambda^\star, \mu^\star)$ satisfies:

$$
\begin{aligned}
& f(x^\star) \leq 0,\ h(x^\star) = 0 && : \text{ primal feasibility} \\
& \lambda^\star \geq 0 && : \text{ dual feasibility} \\
& \lambda_i^\star f_i(x^\star) = 0 \quad \forall i = 1, \ldots, m && : \text{ complementary slackness} \\
& \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \mu_i^\star \nabla h_i(x^\star) = 0 && : \text{ stationarity}
\end{aligned}
$$

Proof: Primal and dual feasibility follow immediately from optimality. Complementary slackness was derived on the previous slide. Stationary holds because $x^\star$ minimizes $L(x, \lambda^\star, \mu^\star)$.

KKT Conditions for Convex Problems

Theorem (Sufficiency): If P is convex and $(x^\star, \lambda^\star, \mu^\star)$ satisfies the KKT conditions, then, $x^\star$ solves P and $(\lambda^\star, \mu^\star)$ solves D.

Proof: The KKT conditions imply

$$\begin{aligned} f_0(x^\star) &= L(x^\star, \lambda^\star, \mu^\star) && \text{(complementary slackness)} \\ &= \inf_x L(x, \lambda^\star, \mu^\star) && \text{(stationarity and convexity of } L \text{ in } x\text{)} \\ &= g(\lambda^\star, \mu^\star). \end{aligned}$$

Thus, $x^\star$ and $(\lambda^\star, \mu^\star)$ are primal and dual feasible with the same objective value. Thus, they are both optimal.

Theorem (Necessity): Assume that a convex P satisfies Slater's condition. If $x^\star$ solves P, there is $(\lambda^\star, \mu^\star)$ such that $(x^\star, \lambda^\star, \mu^\star)$ satisfies the KKT conditions.

Proof: Slater implies strong duality and solvability of D. The claim then follows from the KKT theorem (see previous slide).

Who invented it?

William Karush

Harold W. Kuhn

Albert W. Tucker

The KKT conditions were first named after Kuhn and Tucker, two famous professors from Princeton, who published them in 1951.

Later it was discovered that the conditions had been stated in the MSc thesis of the student William Karush in 1939.

Relevance of the KKT Conditions

The KKT conditions play an important role in optimization.

- Many *algorithms* for convex optimization are conceived as methods for solving the KKT conditions.
- Sometimes it is possible to solve the KKT conditions (and thus, the optimization problem) *analytically*.

Example:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^\top Px + q^\top x + r \\ \text{subject to} & Ax = b \end{array} \qquad (P \in \mathbb{S}^n_+)$$

KKT conditions: $Ax^\star = b, \quad Px^\star + q + A^\top \mu^\star = 0$

$$\iff \quad \begin{pmatrix} P & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^\star \\ \mu^\star \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

Separable Problems

Problem with a *separable objective* and a single constraint:

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & a^\top x = b \end{array}$$

The Lagrangian is also separable:

$$L(x,\mu) = \sum_{i=1}^n f_i(x_i) + \mu(b - a^\top x) = b\mu + \sum_{i=1}^n (f_i(x_i) - \mu a_i x_i)$$

The dual objective is also separable:

$$\begin{aligned} g(\mu) &= b\mu + \inf_x \sum_{i=1}^n (f_i(x_i) - \mu a_i x_i) \\ &= b\mu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) - \mu a_i x_i) \\ &= b\mu - \sum_{i=1}^n f_i^*(\mu a_i) \end{aligned}$$

D is a scalar optimization problem and thus simple. Given $\mu^\star$, we solve the problems $\inf_{x_i} (f_i(x_i) - \mu^\star a_i x_i)$ *separately* to find $x_i^\star$.

Main Take-Away Points

- **KKT conditions:** primal feasibility; dual feasibility; complementary slackness; stationarity
- **KKT theorems:** if Slater's condition holds, then $(x^\star, \lambda^\star, \mu^\star)$ satisfies the KKT conditions iff $x^\star$ and $(\lambda^\star, \mu^\star)$ are primal and dual optimal, respectively