Chapter 5: Optimization for Data Science
Optimality conditions

TANN Chantara

Department of Applied Mathematics and Statistics
Institute of Technology of Cambodia

October 7, 2022

Table of Contents

Optimization problem in standard form

$$
\mathrm{P}: \quad \begin{cases} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ \text{s. t.} & f_i(x) \leq 0 \quad \forall i = 1, \dots, m \\ & h_i(x) = 0 \quad \forall i = 1, \dots, p \end{cases}
$$

Lagrangian

$$
L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)
$$

Primal/dual pair

Primal program: $\inf \mathrm{P} = \inf\limits_{x \in \mathbb{R}^n} \sup\limits_{\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p} L(x, \lambda, \mu) = \inf\limits_{x \in \mathbb{R}^n} f(x)$

Dual program: $\sup \mathrm{D} = \sup\limits_{\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p} \inf\limits_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \sup\limits_{\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p} g(\lambda, \mu)$

Complementary slackness

Assume that strong duality holds ($\inf \mathrm{P} = \sup \mathrm{D}$) and that $x^\star$ is optimal in P while $(\lambda^\star, \mu^\star)$ are 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. Therefore:

- $x^\star$ 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,

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

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, $(\lambda^\star, \mu^\star)$ are optimal in D and strong duality holds (but P might be nonconvex). Then, $(x^\star, \lambda^\star, \mu^\star)$ satisfies:

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

Proof: Primal and dual feasibility follow immediately from optimality. Complementary slackness was derived on the previous slide. Stationarity hold 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

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

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

KKT conditions for convex problems

Theorem (Necessity): Assume that a convex P with differentiable objective and constraint functions 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 (sketch): Slater implies strong duality and solvability of D (we don't show it here). The claim then follows from the KKT theorem.

Who invented it?

William Karush

Harold W. Kuhn

Albert W. Tucker

The KKT conditions were first named after Kuhn and Tucker, two 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: Consider $P \in \mathbb{S}^n_+$ and the optimization problem

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2} x^\top P x + q^\top x + r \\ \text{s.t.} & Ax = b \end{cases}$$

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

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

Separable problems

Consider a problem with a separable objective and a single constraint

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{s. t.} & a^\top x = b \end{cases}$$

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:

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

The dual D is a scalar optimization problem and thus simple, provided one can solve $\inf_{x_i}\{f_i(x_i) - \mu a_i x_i\}$

Main Key Points

- **KKT conditions:** primal feasibility; dual feasibility; complementary slackness; stationarity
- **KKT theorems:** every KKT point of a convex problem constitutes a primal-dual solution pair; every (primal) solution of a convex problem corresponds to a KKT point