Constrained convex optimization problems

Huanle Xu *

April 25, 2017

1 Constrained Optimization Problems

In this chapter, we aim to minimize the following constrained optimization problems:

$$\begin{aligned} \min_{x} \; & f(x) \\ \text{s.t. } & g_i(x) \le 0, \quad \forall i = 1, 2, \cdots, m_1 \\ & h_j(x) = 0, \quad \forall j = 1, 2, \cdots, m_2 \\ & x \in X, \end{aligned}$$

where $f(x)$ and $g_i(x)$ are convex functions and $X \in \mathbb{R}^n$ is a convex set.

Theorem 5 *Let $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be arbitrary functions, and let X be a non–empty subset of $\mathbb{R}^n$. Consider the problem*

$$\begin{array}{ll} \inf & f(x) \\ \text{subject to} & g_i(x) \le 0 \quad \textit{for } i = 1, \ldots, m, \\ & x \in X, \end{array} \tag{19}$$

and define the **Lagrangian function** $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *by*

$$L(x, u) = f(x) + \sum_{i=1}^{m} u_i g_i(x).$$

Let $\bar{x}$ be a feasible solution to (19), and define $I = \{i \in \{1, \ldots, m\} : g_i(\bar{x}) = 0\}$. Suppose that there exists an $\bar{u} \in \mathbb{R}^m$ such that $\bar{u}_i \ge 0$ for $i = 1, \ldots, m$, $\bar{u}_i = 0$ for $i \notin I$, and

$$\bar{x} = \arg\min_{x \in X} L(x, \bar{u}).$$

Then, $\bar{x}$ is a global minimum of (19).

*Huanle Xu is with the College of Computer Science and Technology, Dongguan University of Technology. E-mail: {xuhl}@dgut.edu.cn.

Proof By definition of $\bar{u}$, we have

$$\begin{aligned} f(\bar{x}) &= f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \\ &= \min \left\{ f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x) : x \in X \right\} \\ &\le \inf \left\{ f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x) : x \in X,\ g_i(x) \le 0 \text{ for } i = 1, \ldots, m \right\} \\ &\le \inf \left\{ f(x) : x \in X,\ g_i(x) \le 0 \text{ for } i = 1, \ldots, m \right\}, \end{aligned}$$

where the first equality follows from the fact that $\bar{u}_i g_i(\bar{x}) = 0$ for $i = 1, \ldots, m$. This completes the proof. □

REMARK: If $f, g_1, \ldots, g_m$ are continuously differentiable and convex, then the Lagrangian function $x \mapsto L(x, \bar{u})$ is continuously differentiable and convex in x. Thus, if $X = \mathbb{R}^n$, then by Proposition 2 of Handout 7, we see that $\bar{x}$ satisfies $\nabla L(\bar{x}, \bar{u}) = \mathbf{0}$, or equivalently,

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) = \mathbf{0}.$$

In other words, the first order optimality conditions are sufficient.

The KKT conditions are often useful in gaining insights into the optimization problem at hand, and sometimes they even suggest simpler algorithms for solving the problem. As an illustration, let us consider the following example:

Example 1 (Power Allocation Optimization in Parallel AWGN Channels) *Consider n parallel additive white Gaussian noise (AWGN) channels. For $i = 1, \ldots, n$, the i–th channel is characterized by the channel power gain $h_i \ge 0$ and the additive Gaussian noise power $\sigma_i > 0$. Let p_i denote the transmit power allocated to the i–th channel, where $i = 1, \ldots, n$. The maximum information rate that can be reliably transmitted over the i–th channel is then given by*

$$r_i = \log_2 \left(1 + \frac{h_i p_i}{\sigma_i} \right) = (\ln 2)^{-1} \ln \left(1 + \frac{h_i p_i}{\sigma_i} \right); \tag{20}$$

see [3]. Given a budget P on the total transmit power over n channels, our goal is to allocate power $p_1, \ldots, p_n$ on each of the n channels such that the sum rate of all the channels is maximized. We are thus led to the following formulation:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{n} \ln \left(1 + \frac{h_i p_i}{\sigma_i} \right) \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} p_i \le P, \\ & p_i \ge 0 \qquad\qquad \textit{for } i = 1, \ldots, n. \end{array} \tag{21}$$

It is easy to verify that the objective function of (21) is concave. Hence, problem (21) is a linearly constrained concave maximization problem. Now, by Theorem 4 and the remark after Theorem 5, every solution $(\bar{p}, \bar{v}) \in \mathbb{R}^n \times \mathbb{R}^{n+1}$ *to the following KKT system will yield an optimal solution* $\bar{p} \in \mathbb{R}^n$ *to problem (21):*

$$
\begin{array}{rcllr}
v_0 - v_i & = & \dfrac{h_i}{h_i p_i + \sigma_i} & \text{for } i = 1, \ldots, n, & (a) \\
v_0 \left(\displaystyle\sum_{i=1}^{n} p_i - P \right) & = & 0, & & (b) \\
v_i p_i & = & 0 & \text{for } i = 1, \ldots, n, & (c) \\
v_i & \geq & 0 & \text{for } i = 0, 1, \ldots, n. & (d)
\end{array}
\qquad (22)
$$

To find a solution to the KKT system (22), we proceed as follows. Without loss of generality, we may assume that $h_i > 0$ *for* $i = 1, \ldots, n$*. Then, we have* $v_0 > v_i \geq 0$ *by (22a) and (22d), which implies that*

$$
p_i = \frac{1}{v_0 - v_i} - \frac{\sigma_i}{h_i} \quad \text{for } i = 1, \ldots, n. \qquad (23)
$$

Now, if $p_i > 0$*, then* $v_i = 0$ *by (22c). On the other hand, if* $p_i = 0$*, then in order to satisfy (23) with some* $v_i \geq 0$*, we must have*

$$
\frac{1}{v_0} - \frac{\sigma_i}{h_i} \leq 0.
$$

Hence, we obtain

$$
p_i = \left(\frac{1}{v_0} - \frac{\sigma_i}{h_i} \right)^+ \quad \text{for } i = 1, \ldots, n. \qquad (24)
$$

Moreover, since $v_0 > 0$*, we have* $\sum_{i=1}^{n} p_i = P$ *by (22b). It follows that*

$$
\sum_{i=1}^{n} \left(\frac{1}{v_0} - \frac{\sigma_i}{h_i} \right)^+ = P.
$$

Just like in the case of LP and CLP, given a nonlinear programming problem $\mathscr{P}$ (the primal problem), we can associate with it a dual problem whose properties are closely related to those of $\mathscr{P}$. It turns out that a very general duality theory can be developed for such a primal–dual pair of problems. In particular, we need to make only very few assumptions on the objective function and constraints. Of course, if the primal problem has certain features, then we can gain more specific information about it from the dual problem.

To begin our investigation, consider the following primal problem:

$$
(P) \qquad \begin{array}{rcll} v_p^* & = & \inf & f(x) \\ & & \text{subject to} & g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m_1, \\ & & & h_j(x) = 0 \quad \text{for } j = 1, \ldots, m_2, \\ & & & x \in X. \end{array}
$$

Here, $f, g_1, \ldots, g_{m_1}, h_1, \ldots, h_{m_2} : \mathbb{R}^n \to \mathbb{R}$ are *arbitrary* functions, and X is an *arbitrary* non–empty subset of $\mathbb{R}^n$. For the sake of brevity, we shall write the first two sets of constraints in (P) as $g(x) \leq \mathbf{0}$ and $h(x) = \mathbf{0}$, where $g : \mathbb{R}^n \to \mathbb{R}^{m_1}$ is given by $g(x) = (g_1(x), \ldots, g_{m_1}(x))$ and $h : \mathbb{R}^n \to \mathbb{R}^{m_2}$ is given by $h(x) = (h_1(x), \ldots, h_{m_2}(x))$.

Now, the **Lagrangian dual problem** associated with (P) is the following problem:

$$
(D) \qquad \begin{array}{rcll} v_d^* & = & \sup & \theta(u, v) \equiv \inf_{x \in X} L(x, u, v) \\ & & \text{subject to} & u \in \mathbb{R}_+^{m_1},\ v \in \mathbb{R}^{m_2}. \end{array}
$$

Here, $L : \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}$ is the Lagrangian function given by

$$
L(x, u, v) = f(x) + \sum_{i=1}^{m_1} u_i g_i(x) + \sum_{j=1}^{m_2} v_j h_j(x) = f(x) + u^T g(x) + v^T h(x). \tag{25}
$$

Observe that the above formulation is reminiscent of the penalty function approach, in the sense that we incorporate the primal constraints $g(x) \leq \mathbf{0}$ and $h(x) = \mathbf{0}$ into the objective function of (D) using the Lagrange multipliers u and v. Also, since the set X is arbitrary, there can be many different Lagrangian dual problems for the same primal problem, depending on which constraints are handled as $g(x) \leq \mathbf{0}$ and $h(x) = \mathbf{0}$, and which constraints are treated by X. However, different choices of the Lagrangian dual problem will in general lead to different outcomes, both in terms of the dual optimal value as well as the computational efforts required to solve the dual problem.

Let us now investigate the relationship between (P) and (D). For any $\bar{x} \in X$ and $(\bar{u}, \bar{v}) \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$, we have

$$
\inf_{x \in X} L(x, \bar{u}, \bar{v}) \leq f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) \leq \sup_{u \geq \mathbf{0},\, v \in \mathbb{R}^{m_2}} L(\bar{x}, u, v). \tag{26}
$$

Theorem 6 (Weak Duality Theorem) *Let $\bar{x}$ be feasible for (P) and $(\bar{u}, \bar{v})$ be feasible for (D). Then, we have $\theta(\bar{u}, \bar{v}) \leq f(\bar{x})$. In particular, if $v_d^* = +\infty$, then (P) has no feasible solution.*

Theorem 8 *Let L be the Lagrangian function defined in (25). Suppose that*

1. *X is a compact convex subset of $\mathbb{R}^n$,*
2. *$(u, v) \mapsto L(x, u, v)$ is continuous and concave on $\mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$ for each $x \in X$, and*
3. *$x \mapsto L(x, u, v)$ is continuous and convex on X for each $(u, v) \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$.*

Then, we have

$$
\sup_{u \geq \mathbf{0},\, v \in \mathbb{R}^{m_2}} \min_{x \in X} L(x, u, v) = \min_{x \in X} \sup_{u \geq \mathbf{0},\, v \in \mathbb{R}^{m_2}} L(x, u, v).
$$