Modern Optimization

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

Outline

Definition 2.1

We are interested in the following general optimization problem

$$\begin{aligned} &\min_{x \in X} f_0(x) \\ &\textit{s.t. } h_i(x) = 0, && 1 \leq i \leq m \\ &\textit{and } g_j(x) \leq 0, && 1 \leq j \leq r. \end{aligned}$$

We let $\Omega = \{x \in \mathbb{R}^d : h_i(x) = 0, \forall 1 \leq i \leq m$ *and* $g_j(x) \leq 0, \forall 1 \leq j \leq r$ *defines the set of feasible points.*

Remark 2.1

The minimization is over a domain X which is usually the intersection of the domain of definition of the objective function and the constraints but may also be *simply* some non-negativity constraints.
We expect $\Omega \subset X$.
Example:

$$\min_{x \in X} f_0(x)$$
$$\text{s.t. } x \in \Omega.$$

Definition 2.2

The **primal (optimal) value** *is the value of the function which is optimal (if it exists) over the feasible set:*

$$p^* := \min_{x \in \Omega \cap X} f_0(x).$$

A point $x^* \in \Omega$ such that $f_0(x^*) = p^*$ is called **primal optimal**.

Remark 2.2

So far, we don't know

- if a primal optimal point exists
- if the primal value is finite

Definition 2.3

Given an objective function f_0 and a set of equality constraints h_i and inequality constraints g_j, we define the associate **Lagrangian** *as*

$$\mathcal{L} : \left\{ \begin{array}{rcl} \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^r & \rightarrow & \mathbb{R} \\ (x, \lambda, \mu) & \mapsto & f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x) \end{array} \right.$$

λ and μ are called the **Lagrange multipliers** *or the* **dual variables**.

Remark 2.3

The Lagrangian can be easily rewritten as

$$\mathcal{L}(x, \lambda, \mu) = f_0(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle,$$

with h the m dimensional vector valued function containing all the h_i's component wise. Similarly g is the function stacking all inequality constraints.

Proposition 2.1

For any feasible point $x \in \Omega \subset X$, the objective function is lower bounded by the Lagrangian:

$$\forall x \in \Omega, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r_+, \mathcal{L}(x, \lambda, \mu) \leq f_0(x).$$

Definition 2.4

The **Lagrange dual function** *is defined as*

$$q : \begin{cases} \mathbb{R}^m \times \mathbb{R}^r & \rightarrow & \mathbb{R} \\ (\lambda, \mu) & \mapsto & \inf_{x \in X} \mathcal{L}(x, \lambda, \mu) \end{cases}$$

Proposition 2.2

The dual function q satisfies

1. q *is concave in* λ *and* μ
2. *For any* λ *and* $\mu \geq 0$, $q(\lambda, \mu) \leq p^*$.

Remark 2.4

There is a difference between equality and inequality constraints in how they are handled in the Lagrange dual function.
We might just not consider equality constraints in the remaining unless clearly expressed.

Definition 2.5

Given the general optimization problem, the **dual problem** *is defined as*

$$\begin{aligned} &\max q(\lambda, \mu) \\ &\textit{s.t. } \mu_j \geq 0, \qquad\qquad 1 \leq j \leq r. \end{aligned}$$

If it exists, the maximal value is denoted d^* and is called **the optimal dual value** *or the optimal of the Lagrange dual problem.*

Remark 2.5

Getting to a dual formulation always follows 3 steps:

1. Create the Lagrangian

$$\mathcal{L}(x, \lambda, \mu) = f_0(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle.$$

2. Create the dual Lagrange function:

$$q(\lambda, \mu) = \min_x \mathcal{L}(x, \lambda, \mu).$$

3. Create the dual problem:

$$d^* = \max_{\mu \geq 0} q(\lambda, \mu)$$

Definition 2.6

The **weak duality** *refers to the fact that primal and dual optimal values are ordered:*

$$d^* \leq p^*.$$

The difference between both optimal value is called the **duality gap***:*

$$\Delta = p^* - d^*.$$

Definition 2.7

If the duality gap is 0*, we talk about* **strong duality***:*

$$d^* = p^*.$$

Remark 2.6

It is important to notice the following:

- In general, for nonlinear nonconvex optimization problem, strong duality is not satisfied.
- If strong duality holds, then solving the dual problem solves the primal problem.
- Strong duality is usually satisfied for convex problems.

Example 2.1

Let us look at some basic examples:

1. Linear optimization problem: $\min_{x \in \mathbb{R}^d} c^T x$ subject to $Ax \leq b$ for a vector $c \in \mathbb{R}^d$, a matrix $A \in \mathbb{R}^n \times n$ and observations $b \in \mathbb{R}^n$.
2. $\min_{x \in \mathbb{R}^d} \|x\|_2^2$ subject to $Ax = b$ for some matrix $A \in \mathbb{R}^{n \times m}$ and observations $b \in \mathbb{R}^n$.

Proposition 2.3

The weak duality can be seen as a consequence of the following **minimax inequality***: for any function $\varphi : X \times Y$, we have*

$$\max_{y \in Y} \min_{x \in X} \varphi(x, y) \leq \min_{x \in X} \max_{y \in Y} \varphi(x, y).$$

Definition 2.8

We say that a problem satisfies **Slater's conditions** *is it is strictly feasible, which means:*

$$\exists x \in X : g_j(x) < 0, 1 \leq j \leq r, \text{ and } h_i(x) = 0, 1 \leq i \leq m.$$

Remark 2.7

The **weak Slater** condition is one in which the strict feasibility is replaced by feasibility in case the associated inequality constraint is linear.

Theorem 2.1

If the primal problem is convex and satisfies the weak Slater's conditions the the strong duality holds and

$$\Delta = p^* - d^* = 0.$$

Corollary 2.1

Let f_0 be a convex quadratic function and assume the constraints (equality and inequality) are all linear. The, provided the primal or the dual problem is feasible, strong duality holds.
In particular, strong duality holds for any feasible linear optimization problem.

Example 2.2

Show that the minimizer of the following problem can be found by solving the dual optimization problem:

$$\min_{x \in \mathbb{R}^3} \|x\|_2^2, \qquad \text{subject to } x_1 + x_2 + x_3 \leq -18.$$

Remark 2.8

Convexity is not sufficient for strong duality!

$$\min_{x\in\mathbb{R},y>0} e^{-x}$$

$$\text{s.t. and}\,\frac{x^2}{y} > 0$$