Introduction to nonlinear programming duality

杨俊锋

南京大学数学系

December 28, 2014

Outline[1]

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

[1] Reference: *Convex Optimization* by S. Boyd and L. Vandenberghe.

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

Notation

- The domain of a function $f : R^n \to R$ is denoted by **dom** f.
- Sometimes it is convenient to consider extended real valued function, i.e., $f : R^n \to \bar{R}$, where

$$\bar{R} = R \cup \{+\infty, -\infty\}.$$

Notation

- The domain of a function $f : R^n \to R$ is denoted by **dom** f.
- Sometimes it is convenient to consider extended real valued function, i.e., $f : R^n \to \bar{R}$, where

$$\bar{R} = R \cup \{+\infty, -\infty\}.$$

Consider nonlinear programming (NLP) of the form

$$\begin{aligned} p^* := \min \quad & f_0(x) \\ s.t. \quad & f_i(x) \le 0, \qquad i = 1, 2, \ldots, m, \\ & h_i(x) = 0, \qquad i = 1, 2, \ldots, p, \end{aligned}$$

where f_i, $i = 0, 1, \ldots, m$, and h_i, $i = 1, 2, \ldots, p$, are continuously differentiable functions from R^n to R. Assume that

$$\mathcal{D} = (\cap_{i=0}^{m} \mathbf{dom}\ f_i) \bigcap (\cap_{i=1}^{p} \mathbf{dom}\ h_i) \neq \emptyset.$$

The Lagrange function

The Lagrange function associated with NLP is defined as

$$\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

where $\lambda_i \in R$ $(i = 1, 2, \ldots, m)$ and $\nu_i \in R$ $(i = 1, 2, \ldots, p)$ are, respectively, referred to as the Lagrange multipliers or dual variables associated with $f_i(x) \leq 0$ and $h_i(x) = 0$. The domain of $\mathcal{L}$ is given by

$$\textbf{dom } \mathcal{L} = \mathcal{D} \times R^m \times R^p.$$

The Lagrange dual function

The Lagrange dual function is defined as the minimum value of the Lagrange function over x: for $\lambda \in R^m$, $\nu \in R^p$,

$$
\begin{aligned}
g(\lambda, \nu) &:= \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).
\end{aligned}
$$

The Lagrange dual function

The Lagrange dual function is defined as the minimum value of the Lagrange function over x: for $\lambda \in R^m$, $\nu \in R^p$,

$$\begin{aligned} g(\lambda, \nu) &:= \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right). \end{aligned}$$

- If $\mathcal{L}(x, \lambda, \mu)$ is unbounded below in x, $g(\lambda, \nu)$ takes on the value $-\infty$.

The Lagrange dual function

The Lagrange dual function is defined as the minimum value of the Lagrange function over x: for $\lambda \in R^m$, $\nu \in R^p$,

$$\begin{aligned} g(\lambda, \nu) &:= \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right). \end{aligned}$$

- If $\mathcal{L}(x, \lambda, \mu)$ is unbounded below in x, $g(\lambda, \nu)$ takes on the value $-\infty$.
- $g(\lambda, \nu)$ is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when NLP is not convex.

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* because

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* because

- for any $\lambda \geq 0$ (i.e., each $\lambda_i \geq 0$), any $\nu \in R^p$ and any feasible point $\tilde{x}$, it holds that

$$\mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* because

- for any $\lambda \geq 0$ (i.e., each $\lambda_i \geq 0$), any $\nu \in R^p$ and any feasible point $\tilde{x}$, it holds that

$$\mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Therefore, we have

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* because

- for any $\lambda \geq 0$ (i.e., each $\lambda_i \geq 0$), any $\nu \in R^p$ and any feasible point $\tilde{x}$, it holds that

$$\mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

  Therefore, we have

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

- since this holds for any feasible point $\tilde{x}$, by taking infimum on the right hand side over all feasible points, it follows that

$$g(\lambda, \nu) \leq p^*, \quad \forall (\lambda, \nu) \in (R_+^m, R^p).$$

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* because

- for any $\lambda \geq 0$ (i.e., each $\lambda_i \geq 0$), any $\nu \in R^p$ and any feasible point $\tilde{x}$, it holds that

$$\mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

  Therefore, we have

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

- since this holds for any feasible point $\tilde{x}$, by taking infimum on the right hand side over all feasible points, it follows that

$$g(\lambda, \nu) \leq p^*, \quad \forall (\lambda, \nu) \in (R_+^m, R^p).$$

For $\lambda \geq 0$ and $\nu \in R^p$, the dual function gives a nontrivial lower bound on p^* if $g(\lambda, \nu) \neq -\infty$.

Example – Least-squares solution of linear equations

Let $A \in R^{p \times n}$, where $p < n$. We consider the problem

$$\begin{aligned} \min \quad & \|x\|^2 \\ s.t. \quad & Ax = b. \end{aligned}$$

Example – Least-squares solution of linear equations

Let $A \in R^{p\times n}$, where $p < n$. We consider the problem

$$\begin{aligned} \min \quad & \|x\|^2 \\ s.t. \quad & Ax = b. \end{aligned}$$

- The Lagrange function is give by

$$\mathcal{L}(x,\nu) = \|x\|^2 + \nu^T(Ax - b).$$

- The Lagrange dual function is $g(\nu) = \inf_{x\in R^n} \mathcal{L}(x,\nu)$. Clearly, the infimum is attained at $x = -\frac{1}{2}A^T\nu$. Thus,

$$g(\nu) = -\frac{1}{4}\nu^T AA^T\nu - b^T\nu.$$

- The lower bound property implies that

$$-\frac{1}{4}\nu^T AA^T\nu - b^T\nu \leq \inf\{\|x\|^2 \mid Ax = b\}, \quad \forall \nu \in R^p.$$

Example – Two-way partitioning problem

Let $W \in S^n$. We consider the problem

$$\begin{aligned} p^* = \min \quad & x^T W x \quad \left(= \sum_{i,j=1}^{n} w_{ij} x_i x_j\right) \\ s.t. \quad & x_i^2 = 1, \quad i = 1, 2, \ldots, n. \end{aligned}$$

Example – Two-way partitioning problem

Let $W \in S^n$. We consider the problem

$$p^* = \min \quad x^T W x \quad (= \sum_{i,j=1}^{n} w_{ij} x_i x_j)$$
$$s.t. \quad x_i^2 = 1, \quad i = 1, 2, \ldots, n.$$

- The feasible set contains 2^n points since $x_i \in \{+1, -1\}$. In principle, can be solved by enumeration. However, very difficult for large n, say grater than 50.

Example – Two-way partitioning problem

Let $W \in S^n$. We consider the problem

$$p^* = \min \quad x^T W x \quad (= \sum_{i,j=1}^{n} w_{ij} x_i x_j)$$
$$s.t. \quad x_i^2 = 1, \quad i = 1, 2, \ldots, n.$$

- The feasible set contains 2^n points since $x_i \in \{+1, -1\}$. In principle, can be solved by enumeration. However, very difficult for large n, say grater than 50.
- This problem can be interpreted as a two-way partitioning problem on $\{1, 2, \ldots, n\}$: A feasible x corresponds to a partition

$$\{1, 2, \ldots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}.$$

w_{ij} is the cost of having i and j in the same partition, and $-w_{ij}$ is that of having i and j in different partitions.

- The Lagrange function

$$\mathcal{L}(x,\nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \nu.$$

- The Lagrange dual function

$$g(\nu) = \inf_{x \in R^n} \mathcal{L}(x,\nu) = \begin{cases} -\mathbf{1}^T \nu, & \text{if } W + \mathrm{diag}(\nu) \succeq 0; \\ -\infty, & \text{otherwise.} \end{cases}$$

- Take $\nu = -\lambda_{\min}(W)\mathbf{1}$, we get a lower bound on the optimal value p^* of the difficult combinatorial problem[2]:

$$n\lambda_{\min}(W) \le p^*.$$

[2]This lower bound can also be obtained by relaxing the constraints in the original problem into $\sum_{i=1}^n x_i^2 = n$.

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

The Lagrange dual problem

For each pair $(\lambda, \nu) \in (R^m_+, R^p)$, $g(\lambda, \nu)$ gives a lower bound on the optimal value p^*. A natural question is: What is the best lower bound that can be obtained from $g(\lambda, \nu)$. This leads to the optimization problem (Dual-NLP)

$$\begin{aligned} d^* := \max_{\lambda \in R^m, \nu \in R^p} \quad & g(\lambda, \nu) \\ s.t. \quad & \lambda \geq 0, \end{aligned}$$

which is called the Lagrange dual problem. In this context, the original problem is sometimes called the primal problem.

The Lagrange dual problem

For each pair $(\lambda, \nu) \in (R^m_+, R^p)$, $g(\lambda, \nu)$ gives a lower bound on the optimal value p^*. A natural question is: What is the best lower bound that can be obtained from $g(\lambda, \nu)$. This leads to the optimization problem (Dual-NLP)

$$\begin{aligned} d^* := \max_{\lambda \in R^m, \nu \in R^p} \quad & g(\lambda, \nu) \\ s.t. \quad & \lambda \geq 0, \end{aligned}$$

which is called the Lagrange dual problem. In this context, the original problem is sometimes called the primal problem.

The Lagrange dual problem is a convex optimization problem, since the objective to be maximized is concave and the constraint is convex. This is the case whether or not the primal problem is convex.

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

Weak duality

Theorem (Weak duality)

Suppose x is primal feasible, i.e., $f_i(x) \leq 0$ ($i = 1, \ldots, m$) and $h_i(x) = 0$ ($i = 1, \ldots, p$), and $(\lambda, \nu) \in (R_+^m, R^p)$, i.e., (λ, ν) is dual feasible. Then, it holds that $g(\lambda, \nu) \leq f_0(x)$. Thus, $d^* \leq p^*$.

Weak duality

Theorem (Weak duality)

Suppose x is primal feasible, i.e., $f_i(x) \leq 0$ ($i = 1, \ldots, m$) and $h_i(x) = 0$ ($i = 1, \ldots, p$), and $(\lambda, \nu) \in (R_+^m, R^p)$, i.e., (λ, ν) is dual feasible. Then, it holds that $g(\lambda, \nu) \leq f_0(x)$. Thus, $d^* \leq p^*$.

- If the primal problem is unbounded below, i.e., $p^* = -\infty$, then $d^* = -\infty$, i.e., the dual problem is infeasible.

Weak duality

Theorem (Weak duality)

Suppose x is primal feasible, i.e., $f_i(x) \leq 0$ ($i = 1, \dots, m$) and $h_i(x) = 0$ ($i = 1, \dots, p$), and $(\lambda, \nu) \in (R_+^m, R^p)$, i.e., (λ, ν) is dual feasible. Then, it holds that $g(\lambda, \nu) \leq f_0(x)$. Thus, $d^* \leq p^*$.

- If the primal problem is unbounded below, i.e., $p^* = -\infty$, then $d^* = -\infty$, i.e., the dual problem is infeasible.
- If the dual problem is unbounded above, i.e., $d^* = +\infty$, then $p^* = +\infty$, i.e., the primal problem is infeasible.

Weak duality

Theorem (Weak duality)

Suppose x is primal feasible, i.e., $f_i(x) \leq 0$ ($i = 1, \ldots, m$) and $h_i(x) = 0$ ($i = 1, \ldots, p$), and $(\lambda, \nu) \in (R^m_+, R^p)$, i.e., (λ, ν) is dual feasible. Then, it holds that $g(\lambda, \nu) \leq f_0(x)$. Thus, $d^* \leq p^*$.

- If the primal problem is unbounded below, i.e., $p^* = -\infty$, then $d^* = -\infty$, i.e., the dual problem is infeasible.
- If the dual problem is unbounded above, i.e., $d^* = +\infty$, then $p^* = +\infty$, i.e., the primal problem is infeasible.
- $p^* - d^*$, which is always nonnegative, is referred to as the optimal duality gap or simply duality gap.

Weak duality

Theorem (Weak duality)

Suppose x is primal feasible, i.e., $f_i(x) \le 0$ ($i = 1, \ldots, m$) and $h_i(x) = 0$ ($i = 1, \ldots, p$), and $(\lambda, \nu) \in (R_+^m, R^p)$, i.e., (λ, ν) is dual feasible. Then, it holds that $g(\lambda, \nu) \le f_0(x)$. Thus, $d^* \le p^*$.

- If the primal problem is unbounded below, i.e., $p^* = -\infty$, then $d^* = -\infty$, i.e., the dual problem is infeasible.
- If the dual problem is unbounded above, i.e., $d^* = +\infty$, then $p^* = +\infty$, i.e., the primal problem is infeasible.
- $p^* - d^*$, which is always nonnegative, is referred to as the optimal duality gap or simply duality gap.
- The dual problem is always convex and sometimes can be solved efficiently, and thus yields a lower bound on p^*. For example, the dual of the two-way partition problem is

$$\max_{\nu \in R^n} \{-\mathbf{1}^T \nu : \ s.t. \ W + \operatorname{diag}(\nu) \succeq 0\},$$

which can be solved efficiently even for relatively large n.

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

Strong duality

If the equality $d^* = p^*$ holds, then we say that strong duality holds. In this case the best bound obtained from the Lagrange dual function is tight.

Strong duality

If the equality $d^* = p^*$ holds, then we say that strong duality holds. In this case the best bound obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold.

Strong duality

If the equality $d^* = p^*$ holds, then we say that strong duality holds. In this case the best bound obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold.

Let $f_0, \dots, f_m$ be convex functions. Convex optimization (COP) takes the form

$$
\begin{aligned}
\min \quad & f_0(x) \\
s.t. \quad & f_i(x) \le 0, \quad i = 1, 2, \dots, m, \\
& Ax = b.
\end{aligned}
$$

Slater's condition: Suppose $f_1, \ldots, f_k$ are affine and $f_{k+1}, \ldots, f_m$ are nonlinear. There exists $\bar{x} \in \textbf{relint } \mathcal{D}$ such that

$$\begin{aligned} f_i(\bar{x}) &\leq 0, \quad i = 1, \ldots, k; \\ f_i(\bar{x}) &< 0, \quad i = k+1, \ldots, m; \\ A\bar{x} &= b, \end{aligned}$$

i.e., besides feasibility, $\bar{x} \in \textbf{relint } \mathcal{D}$ is strictly feasible for nonlinear inequality constraints.

Slater's condition: Suppose $f_1, \ldots, f_k$ are affine and $f_{k+1}, \ldots, f_m$ are nonlinear. There exists $\bar{x} \in \textbf{relint}\ \mathcal{D}$ such that

$$\begin{aligned} f_i(\bar{x}) &\leq 0, \qquad i = 1, \ldots, k; \\ f_i(\bar{x}) &< 0, \quad i = k+1, \ldots, m; \\ A\bar{x} &= b, \end{aligned}$$

i.e., besides feasibility, $\bar{x} \in \textbf{relint}\ \mathcal{D}$ is strictly feasible for nonlinear inequality constraints.

Theorem (Strong duality)

Suppose there exists $\bar{x} \in \textbf{relint}\ \mathcal{D}$ *such that the Slater's condition is satisifed. Then, strong duality holds for COP.*

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

Least-squares solution of linear equations

Let $A \in R^{m\times n}$, where $m < n$. Again, we consider the problem

$$p^* = \min\{\|x\|^2 : \; s.t. \; Ax = b\},$$

and its dual problem

$$d^* = \max_{\nu \in R^m} -\frac{1}{4}\nu^T AA^T \nu - b^T \nu.$$

Least-squares solution of linear equations

Let $A \in R^{m \times n}$, where $m < n$. Again, we consider the problem

$$p^* = \min\{\|x\|^2 : \; s.t.\; Ax = b\},$$

and its dual problem

$$d^* = \max_{\nu \in R^m} -\frac{1}{4}\nu^T AA^T \nu - b^T \nu.$$

Slater's condition for this problem is simply that the primal problem is feasible, i.e., $b \in \text{Range}(A)$, in which case

$$p^* < +\infty \quad \text{and} \quad p^* = d^*.$$

In the case $b \notin \text{Range}(A)$, $p^* = +\infty$. By separating hyperplane theorem, there exists $0 \neq z \in R^m$ such that

$$A^T z = 0 \quad \text{and} \quad b^T z < 0.$$

Thus, $d^* = +\infty$. In all, strong duality holds no matter what.

Linear Programming

- Consider linear programming problem in any form.

Linear Programming

- Consider linear programming problem in any form.
- Since all constraints are linear, Slater's condition is simply feasibility.

Linear Programming

- Consider linear programming problem in any form.
- Since all constraints are linear, Slater's condition is simply feasibility.
- Therefore, if the primal problem is feasible, then $p^* < +\infty$ and strong duality holds.

Linear Programming

- Consider linear programming problem in any form.
- Since all constraints are linear, Slater's condition is simply feasibility.
- Therefore, if the primal problem is feasible, then $p^* < +\infty$ and strong duality holds.
- Likewise, if dual LP is feasible, then $d^* > -\infty$ and strong duality holds.

Linear Programming

- Consider linear programming problem in any form.
- Since all constraints are linear, Slater's condition is simply feasibility.
- Therefore, if the primal problem is feasible, then $p^* < +\infty$ and strong duality holds.
- Likewise, if dual LP is feasible, then $d^* > -\infty$ and strong duality holds.
- This leaves only one possible situation in which strong duality for LPs can fail: both the primal and the dual are infeasible.

Linear Programming

- Consider linear programming problem in any form.
- Since all constraints are linear, Slater's condition is simply feasibility.
- Therefore, if the primal problem is feasible, then $p^* < +\infty$ and strong duality holds.
- Likewise, if dual LP is feasible, then $d^* > -\infty$ and strong duality holds.
- This leaves only one possible situation in which strong duality for LPs can fail: both the primal and the dual are infeasible.
- This pathological case can, in fact, occur, as we already know.

Minimum volume covering ellipsoid

Let $\mathcal{E} := \{a \in R^n : a^T X a \leq 1\}$ be an ellipsoid. The volume of $\mathcal{E}$ is proportional to $(\det X^{-1})^{1/2}$. Consider the minimum volume covering ellipsoid problem

$$\begin{aligned} p^* = \min \quad & \log\det X^{-1} \\ s.t. \quad & a_i^T X a_i \leq 1, \\ & i = 1, 2, \ldots, m. \end{aligned}$$

Its dual problem is given by[3]

$$\max_{\lambda \in R_+^m} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T \lambda.$$

Slater's condition is simply that there exists $X \in S_{++}^n$ such that $a_i^T X a_i \leq 1$ for $i = 1, 2, \ldots, m$. Clearly this always holds for $X = \epsilon I$ with $\epsilon > 0$ sufficiently small. Thus, strong duality always holds.

[3] because $\log\det(-Y)^{-1} - n = \sup_{X \succ 0} \langle Y, X \rangle - \log\det X^{-1}$.

Outline

The Lagrange dual function

The Lagrange dual problem

Weak duality

Strong duality

Examples

Saddle-point interpretation

Saddle-point interpretation

Recall that the Lagrange function is

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

Saddle-point interpretation

Recall that the Lagrange function is

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

For any $x \in R^n$, it holds that

$$\sup_{\lambda \in R_+^m,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu) = \begin{cases} f_0(x), & \text{if } f_i(x) \leq 0 \text{ and } h_i(x) = 0 \text{ for all } i; \\ +\infty, & \text{otherwise.} \end{cases}$$

Saddle-point interpretation

Recall that the Lagrange function is

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

For any $x \in R^n$, it holds that

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu) = \begin{cases} f_0(x), & \text{if } f_i(x) \leq 0 \text{ and } h_i(x) = 0 \text{ for all } i; \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore,

$$p^* = \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu)$$

Saddle-point interpretation

Recall that the Lagrange function is

$$\mathcal{L}(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

For any $x \in R^n$, it holds that

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x,\lambda,\nu) = \begin{cases} f_0(x), & \text{if } f_i(x) \le 0 \text{ and } h_i(x) = 0 \text{ for all } i; \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} p^* &= \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x,\lambda,\nu) \\ d^* &= \sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x,\lambda,\nu). \end{aligned}$$

The weak duality $d^* \leq p^*$ is simply

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu),$$

which always holds.

The weak duality $d^* \leq p^*$ is simply

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu),$$

which always holds. In fact, we have

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z),$$

for any $f : R^n \times R^m \to R$, and any $W \subset R^n$ and $Z \subset R^m$. This inequality is called the minimax inequality.

The weak duality $d^* \leq p^*$ is simply

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu),$$

which always holds. In fact, we have

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z),$$

for any $f : R^n \times R^m \to R$, and any $W \subset R^n$ and $Z \subset R^m$. This inequality is called the minimax inequality. If equality holds, i.e.,

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z),$$

we say that the minimax equality holds.

The weak duality $d^* \leq p^*$ is simply

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu),$$

which always holds. In fact, we have

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z),$$

for any $f : R^n \times R^m \to R$, and any $W \subset R^n$ and $Z \subset R^m$. This inequality is called the minimax inequality. If equality holds, i.e.,

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z),$$

we say that the minimax equality holds.

The strong duality means

$$\sup_{\lambda \in R^m_+,\ \nu \in R^p} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \sup_{\lambda \in R^m_+,\ \nu \in R^p} \mathcal{L}(x, \lambda, \nu),$$

which does not, in general, holds.

A pair $(\bar{w}, \bar{z}) \in W \times Z$ is called a saddle-point for f (and W and Z) if

$$f(\bar{w}, z) \leq f(\bar{w}, \bar{z}) \leq f(w, \bar{z})$$

for all $w \in W$ and $z \in Z$.

A pair $(\bar{w}, \bar{z}) \in W \times Z$ is called a saddle-point for f (and W and Z) if

$$f(\bar{w}, z) \leq f(\bar{w}, \bar{z}) \leq f(w, \bar{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\bar{w}$ minimizes $f(w, \bar{z})$ over $w \in W$ and $\bar{z}$ maximizes $f(\bar{w}, z)$ over $z \in Z$:

$$\begin{aligned} f(\bar{w}, \bar{z}) &= \inf_{w \in W} f(w, \bar{z}), \\ f(\bar{w}, \bar{z}) &= \sup_{z \in Z} f(\bar{w}, z). \end{aligned}$$

A pair $(\bar{w}, \bar{z}) \in W \times Z$ is called a saddle-point for f (and W and Z) if

$$f(\bar{w}, z) \leq f(\bar{w}, \bar{z}) \leq f(w, \bar{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\bar{w}$ minimizes $f(w, \bar{z})$ over $w \in W$ and $\bar{z}$ maximizes $f(\bar{w}, z)$ over $z \in Z$:

$$\begin{aligned} f(\bar{w}, \bar{z}) &= \inf_{w \in W} f(w, \bar{z}), \\ f(\bar{w}, \bar{z}) &= \sup_{z \in Z} f(\bar{w}, z). \end{aligned}$$

This implies that

$$\begin{aligned} \inf_{w \in W} \sup_{z \in Z} f(w, z) &\leq \sup_{z \in Z} f(\bar{w}, z) \\ &= f(\bar{w}, \bar{z}) \\ &= \inf_{w \in W} f(w, \bar{z}) \leq \sup_{z \in Z} \inf_{w \in W} f(w, z), \end{aligned}$$

i.e., the minimax equality holds.

If x^* and (λ^*, ν^*) are primal and dual optimal solutions for a problem in which strong duality holds, they form a saddle-point for the Lagrange function, i.e.,

$$\mathcal{L}(x^*, \lambda, \nu) \leq \mathcal{L}(x^*, \lambda^*, \nu^*) \leq \mathcal{L}(x, \lambda^*, \nu^*)$$

for all $x \in \mathcal{D}$ and $(\lambda, \nu) \in R_+^m \times R^p$.

von Neumann's Saddle Point Theorem

Theorem (Classical Saddle Point Theorem)

Assume that X and Z are nonempty convex subsets of R^n and R^m, respectively, and $\phi : X \times Z \to R$ is a function such that $\phi(\cdot, z) : X \to R$ is convex and closed (epigraph is closed) for each $z \in Z$, and $\phi(x, \cdot) : Z \to R$ is concave and closed for each $x \in X$. If X and Z are compact, then the set of saddle points of ϕ is nonempty (i.e., the minimax equality holds) and compact.

Proof.

See Proposition 5.5.3 in *Convex Optimization Theory* by Dimitri P. Bertsekas. □