The content is from Nocedal and Wright (2006) and Bertsekas (1999). Topics marked with ** are optional.

1 Duality
Duality is used to develop important algorithms for constrained optimization such as the augmented Lagrangian algorithm. It also provides a tool to analyze the structure of problems. The theory of duality goes beyond nonlinear optimization, including convex nonsmooth optimization and discrete programming problems. Sometimes it leads to problems that are easier to solve than the original problems.

1.1 The Primal Problem
We consider the following problem as the primal problem:

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) \\
\text{s.t.} & \quad c_i(x) \geq 0, \quad i = 1, 2, \ldots, m.
\end{align*}
\]

For simplicity, we only consider the inequality constraints, but our discussion still can be applied for equality constraints \( g_i(x) = 0 \) by including both \( g_i(x) \geq 0 \) and \( -g_i(x) \geq 0 \).

By collecting the constraints in a vector

\[ c(x) := (c_1(x), c_2(x), \ldots, c_m(x))^T, \]

and collecting the corresponding Lagrange multipliers in a vector

\[ \lambda := (\lambda_1, \lambda_2, \ldots, \lambda_m)^T, \]

we can write the Lagrange function as

\[ \mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x). \]

1.2 The Dual Problem
First, we define the dual objective function \( q : \mathbb{R}^m \to \mathbb{R} \) as follows,

\[ q(\lambda) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda). \]

Note that this requires to find the global minimizer (if exists), which could be extremely difficult. However, when \( f \) is convex and \( c_i \) for \( i = 1, 2, \ldots, m \) are concave,
we can show that $L(\cdot, \lambda)$ is a convex function, in which case the computation of $q(\lambda)$ is more practical.

In many problems the infimum is $-\infty$ for some values of $\lambda$. We define the effective domain of $q$ as

$$\mathcal{D} := \{\lambda \in \mathbb{R}^m : q(\lambda) > -\infty\}.$$ 

Finally, we define the dual problem as follows,

$$\max_{\lambda \in \mathbb{R}^m} q(\lambda) \quad \text{s.t.} \quad \lambda \geq 0.$$ (19.1)

We state an important property of the dual objective function $q(\lambda)$, which makes the dual problem to be well-conditioned even though the corresponding primal problem is not.

**Theorem 19.1.** The dual objective function $q(\lambda)$ is concave and its effective domain $\mathcal{D}$ is convex.

**Proof.** For any $\lambda$ and $\lambda'$ in $\mathbb{R}^m$, any $x \in \mathbb{R}^n$, and any $\alpha \in [0, 1]$, we have

$$L(x, (1 - \alpha)\lambda + \alpha\lambda') = (1 - \alpha)L(x, \lambda) + \alpha L(x, \lambda').$$

By taking infimum of both sides and using the result from Lecture 2 that $\inf \{x_k\} + \inf \{y_k\} \leq \inf \{x_k + y_k\}$, and using the definition of the dual function, we have

$$q((1 - \alpha)\lambda + \alpha\lambda') \geq (1 - \alpha)q(\lambda) + \alpha q(\lambda').$$

That is, $q$ is a concave function. If both $\lambda$ and $\lambda'$ belong to $\mathcal{D}$, then the inequality also implies that $q((1 - \alpha)\lambda + \alpha\lambda') > -\infty$, and therefore $(1 - \alpha)\lambda + \alpha\lambda' \in \mathcal{D}$ as well. That is, $\mathcal{D}$ is a convex set. \qed

Note that in the above theorem we did not assume anything about $f$ and $c_i$. The theorem holds regardless of the convexity/concavity of $f$ and $c_i$.

**1.3 Weak Duality**

**Theorem 19.2.** For any primal feasible $\hat{x}$ and dual feasible $\hat{\lambda}$, we have

$$q(\hat{\lambda}) \leq f(\hat{x}).$$

**Proof.**

$$q(\hat{\lambda}) = \inf_x \{f(x) - \hat{\lambda}^T c(x)\} \leq f(\hat{x}) - \hat{\lambda}^T c(\hat{x}) \leq f(\hat{x}).$$

The last inequality was from the fact that $c(\hat{x}) \geq 0$ and $\hat{\lambda} \geq 0$ from feasibility. \qed

The above theorem allows us to use the dual problems to find a lower bound of the primal optimal objective function value, that is,

$$\max_{\lambda \in \mathbb{R}^m, \lambda \geq 0} q(\lambda) \leq \min_{x \in \mathbb{R}^n, c(x) \geq 0} f(x).$$

The difference between the LHS and the RHS is sometimes called as the duality gap.
2 Duality for Convex Optimization

In a particular case of convex minimization, we have the following result showing that there is no duality gap and the primal and dual objectives are the same at optimality.

First, we write a simplified KKT conditions for the case where we have only inequality constraints:

\[
\nabla f(x^*) - \nabla c(x^*) \lambda^* = 0, \\
c(x^*) \geq 0, \\
\lambda^* \geq 0, \\
\lambda_i^* c_i(x^*) = 0, i = 1, 2, \ldots, m.
\]

Here, we have use the notation that \( \nabla c(x^*) \) is the \( n \times m \) matrix defined by \( \nabla c(x^*) = [\nabla c_1(x^*), \nabla c_2(x^*), \ldots, \nabla c_m(x^*)] \).

2.1 Primal \( \Rightarrow \) Dual

The following result shows that for the pair \( (x^*, \lambda^*) \) that satisfies the KKT conditions of the primal problem, \( \lambda^* \) is the solution of the dual problem in case of convex optimization, and there is no duality gap (in other words strong duality holds).

**Theorem 19.3.** Suppose that \( x^* \) is a solution of the primal problem, and that \( f \) and \( -c_i, i = 1, 2, \ldots, m \), are convex on \( \mathbb{R}^n \) and differentiable at \( x^* \). Then any \( \lambda^* \) for which \( (x^*, \lambda^*) \) satisfies the KKT conditions above is a solution of the dual problem. In particular, \( q(\lambda^*) = f(x^*) \).

**Proof.** Suppose that \( (x^*, \lambda^*) \) satisfies the KKT conditions. Then \( \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \). From the given conditions and \( \lambda^* \geq 0 \), we have that \( \mathcal{L}(\cdot, \lambda^*) \) is a convex and differentiable function. Therefore, for any \( x \in \mathbb{R}^n \), we see that

\[
\mathcal{L}(x, \lambda^*) \geq \mathcal{L}(x^*, \lambda^*) + \nabla_x \mathcal{L}(x^*, \lambda^*)^T (x - x^*) = \mathcal{L}(x^*, \lambda^*),
\]

where the last equality used the fact that \( \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \) from the KKT conditions. Therefore,

\[
q(\lambda^*) = \inf_x \mathcal{L}(x, \lambda^*) = \mathcal{L}(x^*, \lambda^*) = f(x^*) - (\lambda^*)^T c(x^*) = f(x^*),
\]

where the last equality is from the complementary slackness condition in the KKT conditions. From Theorem 19.2, we know that \( q(\lambda) \leq f(x^*) \) for all \( \lambda \geq 0 \), and therefore the above expression tells that \( \lambda^* \) maximizes \( q(\lambda) \).

Note that if the functions are continuously differentiable and a constraint qualification (e.g. LICQ) holds at \( x^* \), then an optimal Lagrange multiplier is guaranteed to exist, by FONC.

2.2 Primal \( \Leftarrow \) Dual

Now the question is that by solving the dual problem, if we can find the primal solution.

**Theorem 19.4.** Suppose that \( f \) and \( -c_i, i = 1, 2, \ldots, m \), are convex and continuously differentiable on \( \mathbb{R}^n \). Suppose that \( x^* \) is a solution of the primal problem at which LICQ holds. Suppose that \( \lambda \) solves the dual problem and that the infimum in \( \inf_x \mathcal{L}(x, \lambda) \) is attained at \( \hat{x} \). Then \( f(x^*) = \mathcal{L}(\hat{x}, \lambda) \). If \( \mathcal{L}(\cdot, \lambda) \) is strictly convex, then \( x^* = \hat{x} \).
Proof. From FONC (with LICQ at $x$), there must exist a Lagrange multiplier vector $\lambda^*$ satisfying the KKT conditions. (Then $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ and hence $x^*$ minimizes $\mathcal{L}(\cdot, \lambda^*)$ due to convexity.) Also, from Theorem 19.3, $\lambda^*$ solves the dual problem, so that

$$\mathcal{L}(x^*, \lambda^*) = q(\lambda^*) = q(\hat{\lambda}) = \mathcal{L}(\hat{x}, \hat{\lambda}).$$  \hfill (19.2)

Furthermore, we also see that

$$f(x^*) = \mathcal{L}(x^*, \lambda^*) = \mathcal{L}(\hat{x}, \hat{\lambda}) = q(\hat{\lambda}),$$

where the first equality comes from the complementarity condition. This tells that (without strict convexity) there is no duality gap.

Suppose that $x^* \neq \hat{x}$ for contradiction. Then from the strict convexity of $\mathcal{L}(\cdot, \lambda)$, we have for $x^* \neq \hat{x}$ that

$$\mathcal{L}(x^*, \hat{\lambda}) > \mathcal{L}(\hat{x}, \hat{\lambda}) + \nabla_x \mathcal{L}(\hat{x}, \hat{\lambda})^T (x^* - \hat{x}) = 0.$$  

Together with (19.2), this implies that

$$\mathcal{L}(x^*, \hat{\lambda}) > \mathcal{L}(\hat{x}, \hat{\lambda}) = \mathcal{L}(x^*, \lambda^*).$$

In particular, we have

$$-\hat{\lambda}^T c(x^*) > - (\lambda^*)^T c(x^*) = 0,$$

where the last equality is from the complementarity condition. Since $\hat{\lambda} \geq 0$ and $c(x^*) \geq 0$, this gives a contradiction, and therefore $x^* = \hat{x}$. \hfill $\Box$

Note that if $f(x)$ is strictly convex, or if one of $-c_i(x)$, $i = 1, \ldots, m$, with $\hat{\lambda}_i > 0$ is strictly convex, then $\mathcal{L}(\cdot, \lambda)$ becomes strictly convex.

2.3 Wolfe Dual

In convex cases, we have a convenient way to construct the dual problem known as the Wolfe dual problem:

$$\max_{x, \lambda} \mathcal{L}(x, \lambda)$$

$$s.t. \quad \nabla_x \mathcal{L}(x, \lambda) = 0,$$

$$\lambda \geq 0.$$

Its relation to the dual problem above is quite straightforward.

**Theorem 19.5.** Suppose that $f$ and $-c_i$, $i = 1, 2, \ldots, m$, are convex and continuously differentiable on $\mathbb{R}^n$. Then solving the Wolfe dual is equivalent to solving the dual problem (19.1).

**Proof.** From given conditions, $\mathcal{L}(\cdot, \lambda)$ is convex for any $\lambda \geq 0$, and therefore $\hat{x}$ satisfying $\nabla_x \mathcal{L}(\hat{x}, \lambda) = 0$ minimizes $\mathcal{L}(\cdot, \lambda)$. That is,

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda) = \mathcal{L}(\hat{x}, \lambda).$$

Therefore, the Wolfe dual can be rewritten as

$$\max_\lambda q(\lambda)$$

$$s.t. \quad \lambda \geq 0,$$

which is the dual problem (19.1). \hfill $\Box$
Ex. Consider the following linear program,
\[
\min_{x \in \mathbb{R}^n} c^T x, \quad \text{s.t.} \quad Ax \geq b, \\
\]
with given \(c \in \mathbb{R}^n\), \(A \in \mathbb{R}^{m \times n}\), and \(b \in \mathbb{R}^m\). The dual objective is (with \(\lambda \in \mathbb{R}^m\)),
\[
q(\lambda) = \inf_{x} [c^T x - \lambda^T (Ax - b)] = \inf_{x} [(c - A^T \lambda)^T x + b^T \lambda]
\]
If \(c - A^T \lambda \neq 0\), then the infimum is \(-\infty\). When \(c - A^T \lambda = 0\), then the dual objective is \(q(\lambda) = b^T \lambda\). In maximizing \(q\), we can exclude the case \(c - A^T \lambda \neq 0\) (since in this case \(q(\lambda) = -\infty\)), and therefore we can write the dual problem as follows,
\[
\max_{\lambda \in \mathbb{R}^m} b^T \lambda \quad \text{s.t.} \quad A^T \lambda = c, \; \lambda \geq 0.
\]
The Wolfe dual can be written as
\[
\max_{\lambda \in \mathbb{R}^m, x \in \mathbb{R}^n} c^T x - \lambda^T (Ax - b) \quad \text{s.t.} \quad A^T \lambda = c, \; \lambda \geq 0.
\]
And substituting \(A^T \lambda - c = 0\) into the objective gives the same dual problem.

3 Duality with Both Equality and Inequality Constraints

Consider the problem
\[
\min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t.} \quad g_i(x) = 0, \; i \in \mathcal{E} \\
\quad \quad \quad \quad \quad c_i(x) \geq 0, \; i \in \mathcal{I}.
\]
We can rewrite the problem as,
\[
\min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t.} \quad g_i(x) \geq 0, \; i \in \mathcal{E} \\
\quad \quad \quad \quad \quad -g_i(x) \geq 0, \; i \in \mathcal{E} \\
\quad \quad \quad \quad \quad c_i(x) \geq 0, \; i \in \mathcal{I}.
\]
Collecting the constraints in vectors,
\[
g(x) = (g_1(x), g_2(x), \ldots, g_{|\mathcal{E}|}(x))^T,
\]
and
\[
c(x) = (c_1(x), c_2(x), \ldots, c_{|\mathcal{I}|}(x))^T,
\]
the corresponding dual problem can be written as,
\[
\max_{x \in \mathbb{R}^n} \inf_{x} f(x) - (\lambda^+)^T g(x) - (\lambda^-)^T (-g(x)) - \mu^T c(x) \\
\text{s.t.} \quad \lambda^+ \geq 0, \; \lambda^- \geq 0, \; \mu \geq 0.
\]
That is,
\[
\max_{\lambda^+, \lambda^-, \mu} \inf_{x} f(x) - (\lambda^+ - \lambda^-)^T g(x) - \mu^T c(x) \\
\text{s.t.} \quad \lambda^+ \geq 0, \; \lambda^- \geq 0, \; \mu \geq 0.
\]
Replacing \(\lambda = \lambda^+ - \lambda^- \in \mathbb{R}^{|\mathcal{E}|}\), we have
\[
\max_{\lambda, \mu} \inf_{x} f(x) - \lambda^T g(x) - \mu^T c(x) \\
\text{s.t.} \quad \lambda \in \mathbb{R}^{|\mathcal{E}|}, \; \mu \geq 0.
\]
And the theory we discussed above generalizes accordingly.
References
