5.3 Lagrangian duality

Consider

\[
(P) \quad \begin{cases} 
\min f(x) \\
\text{s.t. } g_i(x) \leq 0 \quad \forall i \in I \\
x \in X \subseteq \mathbb{R}^n 
\end{cases}
\]

To any Nonlinear Program (NLP) with minimization we can associate a NLP with maximization such that, under some assumptions, the objective function values of the optimal solutions of the two problems coincide. It is possible to solve the primal problem \((P)\) indirectly, by solving the second (dual) problem.

To try to solve \((P)\), we can look for a saddle point of the Lagrange function

\[
L(x, u) = f(x) + \sum_{i \in I} u_i g_i(x) \quad \forall x \in X \quad \forall u \geq 0
\]

**Definition:** (dual function)

\[
w(u) := \inf_{x \in X} L(x, u) \quad \forall u \geq 0
\]

with \(\inf\) instead of \(\min\) because the minimum may not be attained, \(w(u)\) can be equal to \(-\infty\).
Assumptions:

\[ \forall u \geq 0 \text{ such that } w(u) < \infty \]

\[ \exists \bar{x} \in X \text{ such that } w(u) = L(\bar{x}, u) \]

Then

\[ w(u) = \min_{x \in X} L(x, u) \quad \forall u \geq 0 \]

True, for instance, if \( f \) and the \( g_i \)'s are continuous and \( X \) is compact.

**Search for a saddle point** (if it exists):

Dual problem:

\[
(D) \quad \begin{cases} 
\max w(u) \\
\min \quad u \geq 0 
\end{cases}
\]

N.B.: The dual function \( w(u) \) and the dual problem \((D)\) are defined even if no saddle point exists.

Observations:

1) Given a problem \((P)\), we can define different Lagrangian duals depending on which constraints \( g_i(x) \leq 0 \) are "dualized". This choice affects the optimal value of the dual \((D)\) and the computational complexity to evaluate \( w(u) \).

2) The Lagrangian dual is useful to solve large-scale linear programs, convex and non-convex optimization problems, and also discrete optimization problems.
Geometric interpretation:

\[ G = \{(y, z) : y = g(x), z = f(x) \text{ for } x \in X\} \] ...
**Theorem:** (Weak duality)

For all feasible $\bar{x}$ of $(P)$ and $u \geq 0$, we have $w(u) \leq f(\bar{x})$.

**Proof:** By definition of $w(u)$:

$$w(u) \leq f(\bar{x}) + u^T g(\bar{x}) \quad \forall \bar{x} \in X, \forall u \geq 0$$

If $g(\bar{x}) \leq 0$ then $w(u) \leq f(\bar{x})$ because $u \geq 0$.

In particular, the value $w(u)$ for any feasible solution $u \geq 0$ of the dual $(D)$ is a lower bound on the value $f(\bar{x}^*)$ of an optimal solution $\bar{x}^*$ of the primal $(P)$.

**Consequence:**

If a feasible solution $\bar{x}$ of $(P)$ and $\bar{u} \geq 0$ satisfy $w(\bar{u}) = f(\bar{x})$ then $\bar{x}$ is an optimal solution of $(P)$ and $\bar{u}$ is an optimal solution of $(D)$.

For Linear Programming problems the objective function values of optimal solutions of the primal and the dual coincide, for Nonlinear Programs this is not always the case.
**Theorem:** (Strong duality)

i) If \((P)\) has a saddle point \((\bar{x}, \bar{u})\), then

\[
\begin{align*}
\max_{u \geq 0} w(u) &= w(\bar{u}) = f(\bar{x}) = \min \{ f(x) : g(x) \leq 0, \ x \in X \}.
\end{align*}
\]

ii) Conversely, if there exists a feasible solution \(\bar{x}\) of \((P)\) and \(\bar{u} \geq 0\) such that \(w(\bar{u}) = f(\bar{x})\), then \((\bar{x}, \bar{u})\) is a saddle point of \(L(\bar{x}, \bar{u})\).

(According to weak duality, \(\bar{x}\) is an optimal solution of \((P)\) and \(\bar{u}\) is an optimal solution of \((D)\)).

**Proof:**

i) Since \((\bar{x}, \bar{u})\) is a saddle point \(w(\bar{u}) = L(\bar{x}, \bar{u}) = \min_{x \in X} L(\bar{x}, \bar{u})\)

and \(L(\bar{x}, \bar{u}) = f(\bar{x}) + \bar{u}^T g(\bar{x}) = f(\bar{x}) = \min \{ f(x) : g(x) \leq 0, \ x \in X \} \).

Thus \(w(\bar{u}) = f(\bar{x})\).

Because of weak duality \(w(u) \leq f(\bar{x}) \ \forall u \geq 0\), we have \(w(\bar{u}) = \begin{cases} 
\max w(u) \\
\ u \geq 0
\end{cases}\)
ii) If $\overline{x}$ is a feasible solution of $(P)$ and $\overline{u} \geq 0$ such that $w(\overline{u}) = f(\overline{x})$.

By definition of $w(u)$:

$$w(\overline{u}) \leq f(\overline{x}) + \overline{u}^T g(\overline{x}) \quad \forall \overline{x} \in X$$

For $\overline{x} = \overline{x}$, we have $f(\overline{x}) = w(\overline{u}) \leq f(\overline{x}) + \overline{u}^T g(\overline{x})$ and hence $\overline{u}, g_i(\overline{x}) = 0 \quad \forall i \in I$.

Since $\overline{u} \geq 0$, $g_i(\overline{x}) \leq 0 \quad \forall i \in I$ and $w(\overline{u}) = L(\overline{x}, \overline{u}) = \min_{\overline{x} \in X} L(\overline{x}, \overline{u})$ then $(\overline{x}, \overline{u})$ is a saddle point.

**Consequence:**

If $f$ and the $g_i$'s are convex, $X \subseteq R^n$ convex and there exists $a$ such that $g(a) < 0$, then if $(P)$ has a finite optimal solution, there exists a saddle point $(\overline{x}, \overline{u})$ and i) holds, that is

$$\begin{cases} \max_{u \geq 0} w(u) = \min\{f(x) : g(x) \leq 0, x \in X\} \end{cases}$$

N.B.: the optimal values of the objective functions of the two problems coincide, we have **strong duality**.

In general, we can have a **duality gap**:

$$\begin{cases} \max_{u \geq 0} w(u) < \min\{f(x) : g(x) \leq 0, x \in X\} \end{cases}$$
Since we saw that, under certain conditions, it is possible to solve \((P)\) indirectly by solving \((D)\), we need to examine the properties of the dual function.

**Property 1:** The dual function \(w(u)\) is concave.

**Proof:** Take \(u_1, u_2 \geq 0, \lambda \in [0,1]\) and define \(u = \lambda u_1 + (1 - \lambda)u_2\).

By assumption \(\exists \vec{x} \in X\) such that \(w(u) = f(\vec{x}) + u^T g(\vec{x})\).

By definition of \(w(u)\):

\[
\begin{align*}
    w(u_1) &= f(\vec{x}) + u_1^T g(\vec{x}) \\
    w(u_2) &= f(\vec{x}) + u_2^T g(\vec{x})
\end{align*}
\]

\(\Rightarrow \lambda w(u_1) + (1 - \lambda)w(u_2) \leq f(\vec{x}) + u^T g(\vec{x}) = w(u)\).

**Observations:**

- If \(X \subseteq \mathbb{Z}^n\), \(w(u)\) is not continuously differentiable everywhere. Concave piecewise linear function, whose graph is the lower envelope of a finite (or infinite) family of hyperplanes in \(\mathbb{R}^{n+1}\).
- Since \(w(u)\) is concave every local optimum is a global optimum, an ad hoc solution method is needed: the subgradient method.
- \((D)\) is in general easier than \((P)\) -- \(w(u)\) concave and nonnegativity constraints.
Property 2:
For \( \tilde{u} \in \mathbb{R}^{n+} \) let \( X(\tilde{u}) = \left\{ \tilde{x} \in X : f(\tilde{x}) + \tilde{u} \cdot g(\tilde{x}) = w(\tilde{u}) \right\} \) then
\( g(\tilde{x}) \) is a subgradient of \( w(u) \) for each \( \tilde{x} \in X(\tilde{u}) \).

Proof: By definition of \( w(u) \), we have \( w(u) \leq f(\tilde{x}) + u^T g(x) \) \( \forall \tilde{x} \in X \) \( \forall u \geq 0 \).
For \( \tilde{x} \in X(\tilde{u}) \) \( w(u) \leq f(\tilde{x}) + u^T g(\tilde{x}) \)
and by definition of \( X(\tilde{u}) \) \( w(\tilde{u}) \leq f(\tilde{x}) + u^T g(\tilde{x}) \Rightarrow w(u) - w(\tilde{u}) \leq \left( u^T - \tilde{u}^T \right) g(\tilde{x}) \).

Observations:

- It can be shown that every subgradient of \( w(u) \) at \( \tilde{u} \) can be expressed as a convex combination of the subgradients \( g(x) \) with \( \tilde{x} \in X(\tilde{u}) \).
- If \( w \) is continuously differentiable at \( \tilde{u} \) then \( X(\tilde{u}) \) contains a single element and \( g(\tilde{x}) \) coincides with the gradient of \( w(u) \) at \( \tilde{u} \).
Summary

- In general, $(D)$ is easier to solve than $(P)$ -- even if no saddle point exists.

- When a saddle point exists, we can solve $(D)$ instead of $(P)$ and derive, from an optimal solution $u^*$ of $(D)$, an optimal solution $\bar{x}^*$ of $(P)$ by exploiting the characterization of saddle points (that is, by minimizing $L(x, u^*)$ over $X$ and making sure that $g_i(\bar{x}^*) \leq 0$, $\forall i \in I$, and $u^* g_i(\bar{x}^*) = 0$, $\forall i \in I$).

- When no saddle point exists, an optimal solution $u^*$ of $(D)$ provides a lower bound $w(u^*)$ for the optimal value $f(x^*)$ of $(P)$.

- To determine $u^* \geq 0$ that maximizes $w(u)$ we can use the subgradient method that generates a sequence $\{u^i\} \rightarrow u^*$ when $i \rightarrow \infty$.

- For each $u^i$, we have a lower bound $w(u^i)$ for the optimal value $f(x^*)$ of $(P)$ and we determine $\bar{x}^i$ that minimizes $L(x, u^i)$ over $X$. 