Duality Gaps, complementary Slackness & KKT conditions

Recall: Want to solve

\[ \begin{align*}
\text{min } f(x) \\
\text{s.t. } & g_i(x) \leq 0 \quad i = 1, \ldots, m \\
& h_j(x) = 0 \quad j = 1, \ldots, p
\end{align*} \]

So we write the Lagrangian

\[ L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \]

The Lagrangian Dual function is

then \[ F(\lambda, \nu) = \min_{x \in D} L(x, \lambda, \nu) \]

& the dual opt. problem is

\[ \max_{\lambda, \nu} F(\lambda, \nu) \quad \text{s.t. } \lambda \geq 0 \]
Recall: **Weak duality**

\[ \beta^* \leq \lambda^* \quad \text{always holds} \]

(i.e. for all opt. problems)

\[ \max \{ \beta \} \leq \min \{ \lambda \} \quad \text{min of primal problem} \]

\[ \text{Strong duality} \]

\[ \lambda^* = \beta^* \]

Does not always hold, but it sometimes does (e.g., when Slater's conditions hold)

**Duality Gap**

Sometimes solving the dual problem is easier than solving the primal

Even more, since \( \beta^* \leq \lambda^* \)

\[ \max_{(D)} \leq \min_{(P)} \]
Then for any dual-feasible \((\lambda, \nu)\) & primal feasible \(x\):

\[ F(\lambda, \nu) \leq \beta^* \leq \alpha^* \leq f(x) \]

\[ \Rightarrow f(x) - \alpha^* \leq f(x) - F(\lambda, \nu) \]

so if \(f(x) - F(\lambda, \nu)\) is small then we know that \(f(x) - \alpha^*\) is small (i.e. \(x\) is close to being optimal)

**Def:** For an opt. problem with objective \(f\) & dual \(F\), primal feasible \(x\) & dual feasible \((\lambda, \nu)\) the quantity

\[ f(x) - F(\lambda, \nu) \]

is known as the duality gap.
Usefulness: Suppose you have an algorithm that produces a series of primal & dual feasible points $x^{(i)}$ & $(\lambda^{(i)}, \nu^{(i)})$

then if $f(x^{(i)}) - F(\lambda^{(i)}, \nu^{(i)}) < \varepsilon$

we also have $f(x^{(i)}) - x^* < \varepsilon$

Complementary Slackness:

Lemma: Let $x^* \in \mathbb{R}^n$ be primal optimal & let $(\lambda^*, \nu^*)$ be dual optimal & suppose that we have strong duality, then

1. $x^*$ minimizes $L(x, \lambda^*, \nu^*)$

2. $\lambda^*_i \cdot s_i(x^*) = 0 \quad \forall i = 1, \ldots, m$
Prove & Lemma:

(1) \( f(x^*) = F(\lambda^*, \nu^*) \) by strong D.

\[
\begin{align*}
\min_{x} & \quad f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{j=1}^{p} \nu_j^* h_j(x) \\
\text{subject to} & \quad g_i(x) \leq 0, \quad h_j(x) \leq 0 \\
\end{align*}
\]

But (!) \( f(x^*) = f(x^*) \) (of course)

so the inequalities should have all been equalities, i.e.,

\[
\min_{x} \quad f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{j=1}^{p} \nu_j^* h_j(x) = f(x^*)
\]

\( \mu(x, \lambda^*, \nu^*) \)
in order

(which is what we needed \(^\text{a}\) to prove (1))

and

\[
f(x^*) + \sum_{i} x_i^* g_i(x^*) + \sum_{j} z_j^* h_j(x^*) = f(x^*)
\]

so this has to be zero—which proves (2).

Karush-Kuhn-Tucker (KKT) conditions

These are necessary \& sufficient conditions for strong duality

\[
\text{Define } x^* \in \mathbb{R}^n \text{ \& } (\lambda^*, \nu^*) \in \mathbb{R}^m \times \mathbb{R}^p
\]
satisfy the KKT cond's if

\[
\begin{align*}
1. \quad g_i(x^*) &\leq 0 \quad \forall i = 1, \ldots, m \\
2. \quad h_i(x^*) &= 0 \quad \forall i = 1, \ldots, p \\
3. \quad \lambda_i &> 0 \quad \forall i = 1, \ldots, m \\
4. \quad \lambda_i g_i(x^*) &\leq 0 \quad \forall i = 1, \ldots, m
\end{align*}
\]

and

\[
\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0
\]
Theorem

\textbf{Nec. Cond.}

(1) Strong duality \implies

\begin{align*}
\text{Primal optimal pt. } & x^* \in \mathbb{R}^n \\
\text{Dual optimal pts } & (\lambda^*, \nu^*) \in \mathbb{R}^m \times \mathbb{R}^p
\end{align*}

satisfy the KKT conditions

\textbf{Suff. Cond.}

(2) If \( f \) \& \( g_i \) are convex \& if \( h_i \) are all affine \& we have the KKT conditions \implies \text{Strong duality holds}

\textbf{Proof:} (1) The nec cond:

\begin{itemize}
  \item We have 1, 2, 3 by feasibility of \( x^* \) \& \((\lambda^*, \nu^*)\)
  
  \item We have 4 by the Lemma
  
  \item We have 5 since \( x^* \) minimizes \( L(x, \lambda^*, \nu^*) \)
    \begin{align*}
    \Rightarrow \quad \frac{\partial L(x, \lambda^*, \nu^*)}{\partial x} \bigg|_{x^*} = 0
    \end{align*}
\end{itemize}
(2). KKT cond’s (1) & (2) $\Rightarrow x^*$ is primal feasible

- KKT cond (3) $\Rightarrow (\lambda^*, \nu^*)$ is dual feasible
- $f, g_i$ convex & $h_i$ affine

$\Rightarrow$ $L(x, \lambda^*, \nu^*) = f(x) + \sum \lambda^*_i g_i(x) + \sum\nu_i h_i(x)$

is convex in $x$.

- KKT cond (5) $\Rightarrow x^*$ is the global minimizer of $L(x, \lambda^*, \nu^*)$

$\Rightarrow$ $F(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$

$= f(x^*) + \sum \lambda_i^* g_i(x) + \sum\nu_i h_i(x)$

By weak duality

But $F(\lambda^*, \nu^*)$ is weak D

so $(\lambda^*, \nu^*)$ maximizes $F$ & we have strong duality.
We already saw one algorithm for solving constrained opt. problems, namely
Projected gradient descent.

(needed to know how to compute projection)

Another way for solving non-const. opt. problems is to try to (numerically) solve the KKT conditions which is a system of equations & inequalities with

\( n + m + p \) unknowns.

\[ \text{min} \quad f(x) \]
\[ \text{s.t.} \quad g_i(x) \leq 0 \quad i = 1, \ldots, m \]
\[ \Rightarrow L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \]

"Lagrangian algorithm"
\[
\begin{cases}
    x(t+1) = x(t) - \alpha \left( \nabla f(x(t)) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x(t)) \right) \\
    \lambda(t+1) = \Pi^+ (\lambda(t) + \beta_t \nabla g(x(t)))
\end{cases}
\]

where \( \Pi^+ (\lambda) = \tilde{\lambda} \in \mathbb{R}^m \)

with

\[
\begin{cases}
    \tilde{\lambda}_i = \lambda_i & \text{if } \lambda_i \geq 0 \\
    \tilde{\lambda}_i = 0 & \text{if } \lambda_i < 0
\end{cases}
\]

Projection onto the orthant i.e. the set \( \{\lambda : \lambda_i \geq 0, i=1,\ldots,m\} \)

Above, the \( x \) update is a GD step
the \( \lambda \) update is a projected gradient ascent step.

Intuition: want to minimize \( F(x) \)
\& maximize \( F(\lambda) \).
Examples on duality | KKT:

minimize \( x_1^2 + x_2^2 \)

s.t. \[
\begin{cases}
    x_1 + x_2 \geq 4 \\
    x_1 \geq 0 \\
    x_2 \geq 0
\end{cases}
\]

Then \( \mathcal{L}(x, \lambda) = x_1^2 + x_2^2 + \lambda_1(4 - x_1 - x_2) + \lambda_2(-x_1) + \lambda_3(-x_2) \)

\[ F(\lambda) = \min_{x \in D} x_1^2 + x_2^2 + \lambda_1(4 - x_1 - x_2) + \frac{\lambda_2}{2}x_1 + \frac{\lambda_3}{2}x_2 \]

minimizing \[ x_1 = \frac{\lambda_1 + \lambda_2}{2}, \quad x_2 = \frac{\lambda_1 + \lambda_3}{2} \]

\[ F(\lambda) = 4\lambda_1 - \left(\frac{\lambda_1 + \lambda_2}{4}\right)^2 - \left(\frac{\lambda_1 + \lambda_3}{4}\right)^2 \]

\[ \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \]
$F(\lambda)$ is concave & differentiable and we want to maximize it.

Since $\lambda_1, \lambda_2 > 0$, $\lambda_3 > 0$, by inspection we see that to maximize $F(\lambda)$ $\lambda_2^* = 0 \& \lambda_3^* = 0$ (complementary slackness!!)

$\Rightarrow F(\lambda)$ reduces to

$F(\lambda) = 4\lambda_1 - \frac{\lambda_2^2}{2} \Rightarrow \lambda_1^* = 4$

$\Rightarrow \sqrt{F(\lambda^*)} = 8$

On the other hand

$\hat{x}_1 = \frac{\lambda_1^* + \lambda_2^*}{2} = 2$

$\hat{x}_2 = 2$

$\Rightarrow (2, 2)$ is opt.
Notice that at the optimal solution:
\[ x^*_1 + x^*_2 - 4 = 0, \quad \lambda_1^* \neq 0 \]
\[ x^*_1 > 0, \quad \lambda_2^* = 0 \]
\[ x^*_2 > 0, \quad \lambda_3^* = 0 \]
(Complementary slackness)
Exercise: minimize $x_1$

subject to $x_1^2 + x_2^2 = 1$

- What is the sol’n to this opt. problem?
- Formulate the dual.
- Solve it.
- Is $z^* = \beta^*$?