Large-Scale Optimization

L25. DUALITY 3
Duality Gap

For primal feasible \( x \) and dual feasible \((u, v)\),

\[
\text{(duality gap)} := f(x) - g(u, v) \geq 0
\]

For any \( u \geq 0 \) and \( v \), and a primal feasible \( x \)

\[
f(x) \geq f^* \geq \min_{x' \in C} \mathcal{L}(x'; u, v) \geq \min_{x' \in \mathbb{R}^n} \mathcal{L}(x'; u, v) =: g(u, v)
\]

\[
\Rightarrow \quad 0 \leq f(x) - f^* \leq f(x) - g(u, v)
\]

Provides a practical stopping criterion:

- zero duality gap \( \Rightarrow \) \( x \) is primal optimal & \((u, v)\) is dual optimal
- check “duality gap < epsilon” for some epsilon > 0
- due to limited precision, duality gap < 0 can happen in code
# First-Order Optimality Conditions

<table>
<thead>
<tr>
<th>Necessary, under</th>
<th>Sufficient, under</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ (constraint qualification, e.g. LICQ), or strong duality</td>
<td>Convex optimization</td>
</tr>
<tr>
<td>No convex opt assumption</td>
<td></td>
</tr>
</tbody>
</table>

When the problem is a convex opt,

<table>
<thead>
<tr>
<th>Necessary, under</th>
<th>Always sufficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ (constraint qualification, e.g. LICQ), or strong duality (e.g. convex opt + Slater)</td>
<td></td>
</tr>
</tbody>
</table>
FONC (Under Strong Duality)

Let $x^*$ and $(u^*, v^*)$ be primal and dual solutions satisfying strong duality. Then $(x^*, u^*, v^*)$ satisfies the KKT conditions.

First, $x^*$ and $(u^*, v^*)$ are primal and dual feasible.

$$f(x^*) = g(u^*, v^*)$$
$$= \min_{x \in \mathbb{R}^n} \mathcal{L}(x; u^*, v^*)$$
$$\leq \mathcal{L}(x^*; u^*, v^*)$$
$$\leq f(x^*)$$

Therefore, all inequalities should hold as equalities.
FONC (Under Strong Duality)

\[ f(x^*) = g(u^*, v^*) \]
\[ = \min_{x \in \mathbb{R}^n} \mathcal{L}(x; u^*, v^*) \]
\[ = \mathcal{L}(x^*; u^*, v^*) \]
\[ = f(x^*) \]

\[ \mathcal{L}(x^*; u^*, v^*) = f(x^*) - \sum_{i=1}^{m} u^*_i h_i(x^*) - \sum_{j=1}^{n} v^*_j l_j(x^*) \]

\[ \Rightarrow u^*_i h_i(x^*) = 0 \text{ should hold for all } i. \]

No assumption on the convexity of the problem!
Strong Duality: Dual $\rightarrow$ Primal

**Strong duality** allows us to characterize primal solutions from dual solutions.

A result in the proof of “FONC + strong duality”: given a dual solution $(u^*, v^*)$, any primal solution $x^*$ is a solution of $\min_{x \in \mathbb{R}^n} L(x; u^*, v^*)$.

When $L(x; u^*, v^*)$ is convex in $x$, its minimizer ($x^*$) is characterized by

$$0 \in \partial_x L(x^*; u^*, v^*)$$

That is, this can be solved characterizes a primal solution.

If $\min_{x \in \mathbb{R}^n} L(x; u^*, v^*)$ has a unique minimizer, then it must be the unique primal solution.
Example (B & V: Ex 5.4)

$$\min_{x \in \mathbb{R}^n} f_0(x) = \sum_{i=1}^{n} f_i(x_i) \quad \text{s.t. } a^T x = b$$

(Feasible set) \cap \text{ dom } f_0 \text{ is nonempty}

Each $f_i : \mathbb{R} \to \mathbb{R}$ is strictly convex and differentiable

$\Rightarrow$ there exists a unique minimizer $x^*$

Also, Slater’s condition hold $\Rightarrow$ strong duality
\[
\min_{x \in \mathbb{R}^n} f_0(x) = \sum_{i=1}^{n} f_i(x_i) \quad \text{s.t. } a^T x = b
\]

Dual function: \[ g(\lambda) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda) \]

\[
= \inf_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} f_i(x_i) + \lambda (b - a^T x) \right\}
\]

\[
= b\lambda + \sum_{i=1}^{n} \inf_{x_i \in \mathbb{R}} \left\{ f_i(x_i) - a_i \lambda x_i \right\}
\]

\[
= b\lambda - \sum_{i=1}^{n} f_i^*(a_i \lambda) \quad \text{convex in } v
\]

\[ f_i^*(v) := \sup_{x_i \in \mathbb{R}} \{ vx_i - f_i(x_i) \} \]

conjugate of \(f_i\)
Dual problem:

\[ f_i^*(v) := \sup_{x_i \in \mathbb{R}} \{ vx_i - f_i(x_i) \} \]

\[
\max_{\lambda \in \mathbb{R}} g(\lambda) = \max_{\lambda} \left\{ b\lambda - \sum_{i=1}^{n} f_i^*(a_i\lambda) \right\}
\]

- This is a convex optimization (i.e., maximization of a concave obj. function)
- Simple algorithm can solve a convex problem over one scalar variable
- Suppose that we have a dual solution \( \lambda^* \):
  \( \mathcal{L}(x; \lambda^*) \) is strictly convex and thus has a unique minimizer \( \tilde{x} \)
  Strong duality \( \Rightarrow \) \( x^* \) minimizes \( \mathcal{L}(x; \lambda^*) \) \( \Rightarrow \) \( \tilde{x} = x^* \)
- Therefore, we can recover \( x^* \) from \( \nabla_x \mathcal{L}(x; \lambda^*) = 0 \), i.e. by solving equations \( f_i'(x_i^*) = -\lambda^* a_i \)
Today

Conjugation

Biconjugation

Connection to Lagrangian dual
Conjugate

\[ f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \text{ not necessarily convex} \]

\[ f \not\equiv +\infty, \text{ there exists an affine function minorizing } f \text{ on } \mathbb{R}^n \]

\[ \Rightarrow f(x) > -\infty \ \forall x \quad \text{dom } f := \{x : f(x) < +\infty\} \neq \emptyset \]

\[ f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \text{ is the conjugate of } f \text{ defined by} \]

\[ f^*(y) := \sup_{x \in \text{dom } f} \{y^T x - f(x)\} \]

The mapping \( f \mapsto f^* \) is called the conjugacy operation, conjugation, or Legendre-Fenchel transform.
Conjugate: Basic Properties

\[ f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, \text{ not necessarily convex} \]
\[ f \not\equiv +\infty, \quad f(x) \geq y_0^T x - b \quad \forall x \text{ for some } (y_0, b). \]

\[ \Rightarrow f(x) > -\infty \quad \forall x \quad \text{dom } f := \{x : f(x) < +\infty\} \neq \emptyset \]

\[ f^*(y) := \sup_{x \in \text{dom } f} \{y^T x - f(x)\} \]

\[ f^*(y_0) \leq b \]
\[ f^*(y) > -\infty \]

Note that \( f^* \) also satisfies the required conditions for conjugation.
Conjugate

\[ f^*(y) := \sup_{x \in \text{dom } f} \{ y^T x - f(x) \} = -\inf_{x \in \text{dom } f} \{ f(x) - y^T x \} \]

Figure 3.8 A function \( f : \mathbb{R} \to \mathbb{R} \), and a value \( y \in \mathbb{R} \). The conjugate function \( f^*(y) \) is the maximum gap between the linear function \( yx \) and \( f(x) \), as shown by the dashed line in the figure. If \( f \) is differentiable, this occurs at a point \( x \) where \( f'(x) = y \).

B & V, p.91
Why Conjugate?

We know the gradient of a differentiable function $f$ plays a key role.

Consider gradient as a mapping, $x \mapsto y(x) = \nabla f(x)$.

An interesting object is the inverse, $x(\cdot) = (\nabla f)^{-1}$.

Using the conjugate $f^*(y) = y^T x(y) - f(x(y))$,

the inverse becomes another gradient mapping, $x(\cdot) = \nabla f^*$

Conjugate is a simple function having such a property.
Properties of Conjugate I

\[ \alpha \in \mathbb{R} \]

\[ g(x) = f(x) + \alpha \quad \Rightarrow \quad g^*(y) = f^*(y) - \alpha \]

\[ g(x) = \alpha f(x), \ \alpha > 0 \quad \Rightarrow \quad g^*(y) = \alpha f^*(y/\alpha) \]

\[ g(x) = f(\alpha x), \ \alpha \neq 0 \quad \Rightarrow \quad g^*(y) = f^*(y/\alpha) \]

\[ g(x) = f(x - x_0) \quad \Rightarrow \quad g^*(y) = f^*(y) + y^T x_0 \]

\[ g(x) = f(x) + y_0^T x \quad \Rightarrow \quad g^*(y) = f^*(y - y_0) \]
Properties of Conjugate II

\[ f_1 \leq f_2, \; \text{dom} \; f_1 \supseteq \text{dom} \; f_2 \quad \Rightarrow \quad f_1^* \geq f_2^* \]

\[ \text{dom} \; f_1 \cap \text{dom} \; f_2 \neq \emptyset, \; \alpha \in [0, 1], \]

\[ [\alpha f_1 + (1 - \alpha) f_2]^* \leq \alpha f_1^* + (1 - \alpha) f_2^* \]

\[ f(x) = \sum_{j=1}^{m} f_j(x_j) \quad \Rightarrow \quad f^*(y_1, \ldots, y_m) = \sum_{j=1}^{m} f_j^*(y_j) \]
Fenchel’s Inequality

aka Fenchel-Young Inequality: \( \forall (x, y) \in \text{dom } f \times \mathbb{R}^n, \)

\[
f(x) + f^*(y) \geq x^T y.
\]

Obvious from the definition.
Ex. Indicator Function

\[ f(x) = I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{o.w.} \end{cases} \]

Conjugate:

\[ f^*(y) = l_C^*(y) = \sup_{x \in C} y^T x \]

\( f^* \) is called the support function of \( C \)
Ex. Norms

\[ f(x) = \|x\| \quad f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \leq 1 \\ \infty & \text{o.w.} \end{cases} \]

where \( \| \cdot \|_* := \max_{\|z\| \leq 1} z^T y \) is the \textbf{dual norm} of \( \| \cdot \| \)

\[ f^*(y) := \sup_{x \in \mathbb{R}^n} \{ y^T x - \|x\| \} \]

if \( \|y\|_* > 1 \), then there exists \( z \) s.t. \( \|z\| \leq 1 \) and \( z^T y = \|y\|_* > 1 \),

\[ (t z)^T y - \|t z\| = t(z^T y - \|z\|) \to \infty \text{ with } t \to \infty \]

if \( \|y\|_* \leq 1 \), then \( z^T y - \|z\| \leq \|z\| \|y\|_* - \|z\| \leq 0 \)
Biconjugation

Note that $f^*$ satisfies the required conditions for conjugation.

$$f^{**}(x) = (f^*)^*(x) = \sup_{y \in \mathbb{R}^n} \{x^T y - f^*(y)\}$$

$$f^{**} \leq f$$

$$\text{epi } f^{**} = \text{cl conv epi } f$$

If $f$ is convex, $\text{epi } f^{**} = \text{cl epi } f$

If $f$ is convex and closed, $f^{**} = f$
Dual of an Unconstrained Problem

Find the Lagrange dual of an unconstrained problem,

$$\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Consider an equivalent constrained problem:

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^n} f(x) + g(z), \quad \text{s.t.} \quad x = z$$

Note that there are multiple ways to create an equivalent constrained problem, and they will lead to different Lagrange dual problems!
\[
\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^n} f(x) + g(z), \quad \text{s.t.} \quad x = z
\]

Lagrange dual function:

\[
g(u) = \inf_{x \in \mathbb{R}^n, z \in \mathbb{R}^n} \{ f(x) + g(z) + u^T (z - x) \}
\]

\[
= \inf_{x \in \mathbb{R}^n} \{ -u^T x + f(x) \} + \inf_{z \in \mathbb{R}^n} \{ u^T z + g(z) \}
\]

\[
= -\sup_{x \in \mathbb{R}^n} \{ u^T x - f(x) \} - \sup_{z \in \mathbb{R}^n} \{ (-u)^T z - g(z) \}
\]

\[
= -f^*(u) - g^*(-u)
\]

Lagrange dual problem:

\[
\max_{u \in \mathbb{R}^n} -f^*(u) - g^*(-u)
\]
Ex. Dual Problem

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) + I_C(x) \\
\max_{u \in \mathbb{R}^n} & \quad -f^*(u) - I_C^*(-u) \\
\min_{x \in \mathbb{R}^n} & \quad f(x) + \|x\| \\
\max_{u \in \mathbb{R}^n} & \quad -f^*(u) \text{ s.t. } \|u\|_* \leq 1
\end{align*}
\]
Ex. Dual of Lasso

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|^2_2 + \lambda \|x\|_1
\]

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - z\|^2_2 + \lambda \|x\|_1, \text{ s.t. } z = Ax
\]

Dual function:

\[
g(u) = \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^n} \frac{1}{2} \|y - z\|^2 + \lambda \|x\|_1 + u^T(z - Ax)
\]

\[
= \min_{z \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - z\|^2 + u^Tz \right\} + \min_{x \in \mathbb{R}^n} \left\{ \lambda \|x\|_1 - u^TAx \right\}
\]

\[
= \frac{1}{2} \|y\|^2 - \frac{1}{2} \|y - u\|^2 - \lambda \max_{x \in \mathbb{R}^n} \{v^Tx - \|x\|_1\}, \quad v := ATu/\lambda
\]
Dual problem: \[
\max_{u \in \mathbb{R}^n} \frac{1}{2} \|y\|^2 - \frac{1}{2} \|y - u\|^2 \quad \text{s.t.} \quad \|A^T u\|_\infty \leq \lambda
\]

Or, equivalently,

\[
\min_{u \in \mathbb{R}^n} \frac{1}{2} \|y - u\|^2 \quad \text{s.t.} \quad \|A^T u\|_\infty \leq \lambda
\]