Large-Scale Optimization

L22. DUALITY 2
Today

Duality gap
KKT conditions + strong duality
Equivalence of Constrained vs Lagrange formulations
Dual $\rightarrow$ Primal
Since the last time, we’ve using the following definition of Lagrange dual function,

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

In fact, the duality theory works with general definition with inf & sup

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

What is the difference, min vs. inf ?
Lagrange Dual

\[ f^* = \min_{x \in \mathbb{R}^n} f(x) \]

Primal:
\[
\begin{align*}
\text{s.t.} & \quad h_i(x) \geq 0, \ i = 1, \ldots, m \\
& \quad \ell_j(x) = 0, \ j = 1, \ldots, r \\
\end{align*}
\]
\[ C : \text{feasible set} \]

Lagrange function:
\[
\mathcal{L}(x; u, v) = f(x) - \sum_{i=1}^{m} u_i h_i(x) - \sum_{j=1}^{r} v_j \ell_j(x)
\]

For any \( u \geq 0 \) and \( v \),
\[
f^* \geq \min_{x \in C} \mathcal{L}(x; u, v) \geq \min_{x \in \mathbb{R}^n} \mathcal{L}(x; u, v) =: g(u, v)
\]

Lagrange dual function

Lagrange dual problem:
\[
g^* = \max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} g(u, v) \]
\[
\text{s.t.} \quad u \geq 0
\]

Weak duality: \( f^* \geq g^* \)

Strong duality: \( f^* = g^* \)
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
Constrained Optimization

\[ \min_{x \in \mathbb{R}^n} f(x) \]

s.t. \( h_i(x) \geq 0, \quad i = 1, \ldots, m \)

\[ \ell_j(x) = 0, \quad j = 1, \ldots, r \]

C : feasible set

Basic requirements to find a solution:

• C is nonempty
• C is a closed set
KKT Conditions

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

\[0 \in \partial_x L(x; u, v) = \partial f(x) - \sum_{i=1}^{m} u_i \partial h_i(x) - \sum_{j=1}^{r} v_j \partial \ell_j(x)\]

Lagrange optimality / stationary

\[h_i(x) \leq 0, \quad \ell_j(x) = 0 \quad \forall i, j\]

Primal feasibility

\[u_i \geq 0, \quad \forall i\]

Dual feasibility

\[u_i \ h_i(x) = 0, \quad \forall i\]

Complementary slackness
First-Order Necessary Optimality Condition (FONC)

Let \( x^* \) be a (local) minimizer, at which CQ holds. Then there exists Lagrange multipliers \((u^*, v^*)\) satisfying the KKT conditions at \((x^*, u^*, v^*)\).
Constraint Qualification

Required so that Lagrange multipliers exist satisfying the KKT conditions

- **LICQ (Linear independence CQ):** the gradients of active constraints are linearly independent at $x^*$
  - $\rightarrow$ Lagrange multipliers exist and are **unique**

- **MFCQ (Mangasarian-Fromovitz CQ):**
  - there exists $w \in \mathbb{R}^n$ s.t.
    \[
    \nabla h_i(x^*)^T w > 0, \quad \text{for all active inequality constraints}
    \]
    \[
    \nabla \ell_j(x^*)^T w = 0, \quad \text{for all equality constraints,}
    \]
  - and the set of equality constraint gradients is linearly independent.
Constraint Qualification

- LICQ implies MFCQ
- LICQ is one of the most restrictive CQs
- ... other CQs with less restrictive conditions

- Slater’s condition:
  there exists $x \in \text{relint } C$ s.t.
  \[
  \nabla h_i(x) > 0, \text{ for all non-affine inequality constraints}
  \]
  \[
  \nabla \ell_j(x) = 0, \text{ for all equality constraints.}
  \]
  \[\Rightarrow\] then strong duality holds for convex optimization problem
  - This works as a CQ in convex optimization
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.

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

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

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

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

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

No assumption on the convexity of the problem!
Sufficient Optimality Condition (Under Convexity)

Let \( x^* \) and \( (u^*, v^*) \) satisfy the KKT conditions.

Then, the duality gap is zero: \( x^* \) and \( (u^*, v^*) \) are primal and dual solutions.

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

\[
= f(x^*) - \sum_{i=1}^{m} u_i^* h_i(x^*) - \sum_{j=1}^{r} v_j^* \ell_j(x^*)
\]

\[
= f(x^*)
\]

The primal needs to be a convex opt problem: \( f(x) \) convex, \( h_i(x) \) concave, \( \ell_j(x) \) affine

\Rightarrow L(x; u^*, v^*) \) is convex in \( x \)

\Rightarrow 0 \in \partial_x L(x^*; u^*, v^*) \) is sufficient \( x^* \) to be a minimizer of \( L(x; u^*, v^*) \)
First-Order Optimality Conditions

Necessary, under
- CQ (constraint qualification, e.g. LICQ), or strong duality
- No convex opt assumption

Sufficient, under
- Convex optimization

When the problem is a convex opt,

Necessary, under
- CQ (constraint qualification, e.g. LICQ), or strong duality (e.g. convex opt + Slater)

Always sufficient
Why Both Conditions Matter?

For checking optimality:

- **(Sufficiency)** KKT satisfied $\Rightarrow$ a solution

- **KKT not satisfied $\Rightarrow$ not a solution?**
  - This is given by the contrapositive of "(Necessity) a solution $\Rightarrow$ KKT satisfied"
  - This means that we do need CQ or strong duality, to reject non-optimal points even in convex opt
Constrained vs. Lagrange Forms

\[(C) \quad \min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h(x) \leq t \quad h : \mathbb{R}^n \rightarrow \mathbb{R}^m\]

\[(L) \quad \min_{x \in \mathbb{R}^n} f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)\]

The two formulations are often treated as interchangeable, for some appropriate choices of the parameters \(t\) and \(\lambda\).

Q: Are they really equivalent (under the convex opt assumption)?
(C) → (L): suppose that \( x^* \) is a solution of (C) at which CQ or strong duality holds. By FONC, there exist Lagrange multipliers \( \mu_i^* \geq 0 \) satisfying the KKT conditions of (C),

\[
\partial_x \mathcal{L}(x^*; \mu^*) = \nabla f(x^*) + \sum_i \mu_i^* \nabla h_i(x^*) = 0, \quad (1)
\]

\[
h_i(x^*) \leq t_i \quad (2)
\]

\[
\mu_i^* (h_i(x^*) - t_i) = 0 \quad \forall i. \quad (3)
\]

where (1) becomes the optimality condition of \( x^* \) in (L) if we set \( \lambda_i = \mu_i^* \).

(L) → (C): suppose that \( x^* \) is a solution of (L). Obviously, (1) holds when \( \mu_i^* = \lambda_i \). (2) and (3) holds when \( t_i = h_i(x^*) \) for all \( i \). Under the convexity of (C), the KKT conditions (1–3) are sufficient for \( x^* \) to be a solution of (C).
Equivalence?

\[ \bigcup_{\lambda \geq 0} \{ \text{solutions of (L)} \} \subseteq \bigcup_t \{ \text{solutions of (C)} \} \]

\[ \bigcup_{\lambda \geq 0} \{ \text{solutions of (L)} \} \supseteq \bigcup_{t: \text{(C) is strictly feasible}} \{ \text{solutions of (C)} \} \]

or, any CQ

Typical examples: \( h(x) = \|x\|_1 \) and \( h(x) = \|x\|_2^2 \)

- For these, \( t=0 \) is the only value without strict feasibility