Lagrangian Duality

Wing-Kin (Ken) Ma

Department of Electronic Engineering,
The Chinese University of Hong Kong, Hong Kong

Lesson 10, ELEG5481

Recall that...

- Optimization problem in a standard form (not necessarily convex)

$$\begin{aligned} \min \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{P}$$

with problem domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=1}^{p} \mathbf{dom} h_i$.

- We call (P) the **primal problem** (to distinguish from dual introduced soon).

- Primal optimal value

$$p^\star = \inf_{x \in C} f_0(x)$$

where C is the primal feasible set

$$C = \{x \mid f_i(x) \leq 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p\}$$

Lagrangian

- **Lagrangian** of the primal problem (P)

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

  where $\lambda \succeq 0$ (or $\lambda_i \geq 0$ for $i = 1, \ldots, m$), & $\nu \in \mathbf{R}^p$.

- Fixing an x, $L(x, \lambda, \nu)$ is an affine function of (λ, ν) (convex & concave).

- **Lagrange dual function:**

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

  By pointwise infimum of concave functions, g **is concave** (even though (P) is nonconvex).

Linear approximation interpretation of Lagrangian:

- The original problem

$$\begin{aligned} &\min \; f_0(x) \\ &\text{s.t. } f_i(x) \leq 0, \; i = 1, \ldots, m \quad h_i(x) = 0, \; i = 1, \ldots, p \end{aligned}$$

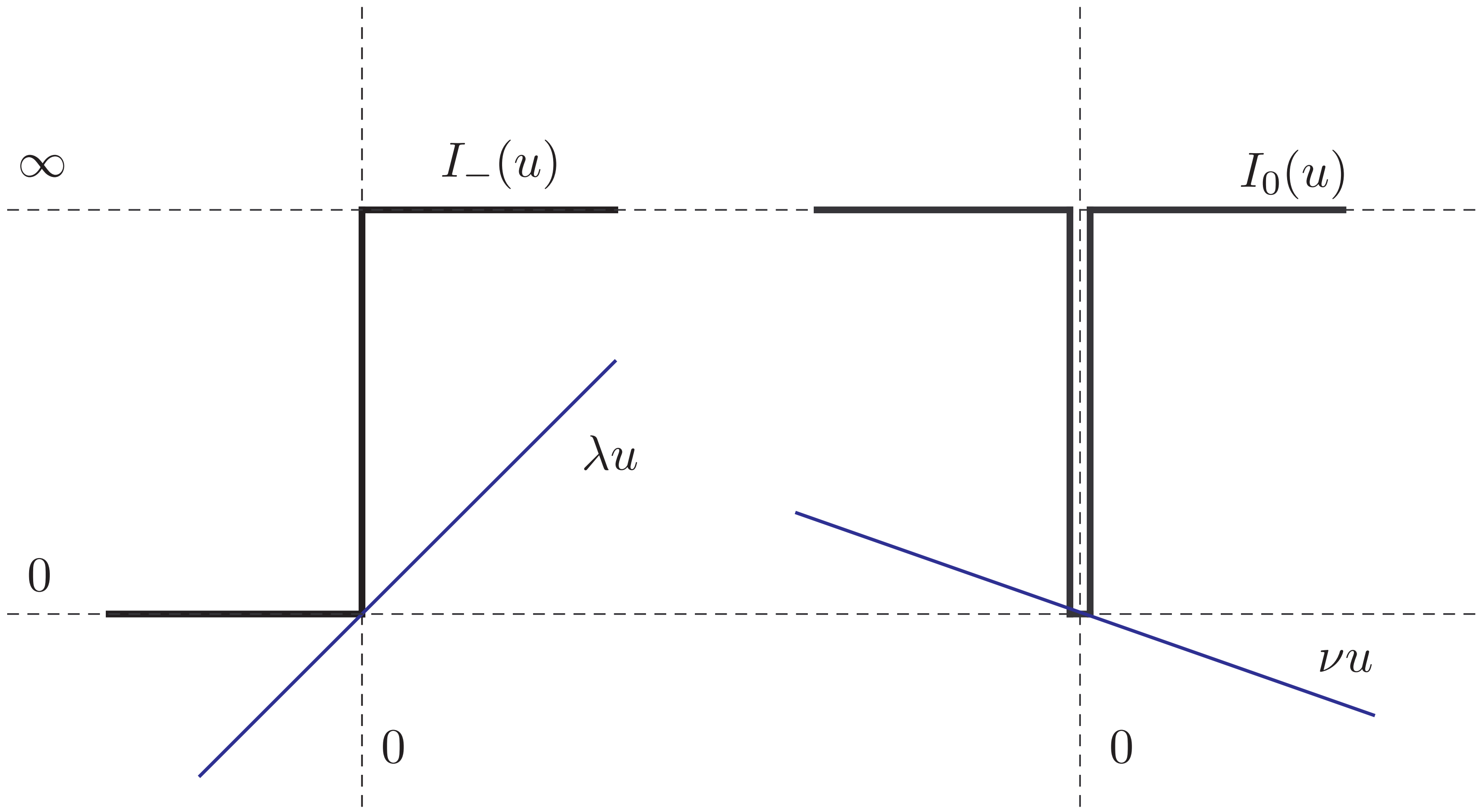
  may be written as an unconstrained problem

$$\min \; f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x))$$

  where $I_-(u) = 0$ for $u \leq 0$, $I_-(u) = \infty$ otherwise; $I_0(u) = 0$ for $u = 0$, $I_0(u) = \infty$ otherwise.

- Lagrangian may be seen as a linear approx. of the reformulated objective

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$


∞
$I_-(u)$
λu
0
0
$I_0(u)$
νu
0

Lower Bound Property and Dual Problem

- **Lower bound on primal optimal:** For any $\lambda \succeq 0$ & ν,

$$g(\lambda, \nu) \leq p^\star$$

- **Dual problem** of (P)

$$\begin{aligned} \max_{\lambda, \nu} \ & g(\lambda, \nu) \\ \text{s.t.} \ & \lambda \succeq 0 \end{aligned} \qquad \text{(D)}$$

  - Motivation: compute the best lower bound on $p^\star$.
  - (D) **is always convex**, whether or not (P) is convex.

- Dual optimal value:

$$d^\star = \sup_{\lambda \succeq 0, \nu} g(\lambda, \nu)$$

Example: Standard form LP

$$\begin{array}{l} \min \ c^T x \\ \text{ s.t. } Ax = b, \quad x \succeq 0 \end{array}$$

Its Lagrangian

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (b - Ax) = (c - \lambda - A^T \nu)^T x + b^T \nu$$

Its dual function

$$g(\lambda, \nu) = \begin{cases} b^T \nu, & c - \lambda - A^T \nu = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Hence its dual problem is

$$\begin{array}{ll} \max & b^T \nu \\ \text{s.t.} & c - A^T \nu = \lambda, \ \lambda \succeq 0 \end{array} \iff \begin{array}{ll} \max & b^T \nu \\ \text{s.t.} & c - A^T \nu \succeq 0 \end{array}$$

which is an LP in inequality form.

Example: Homogenous Boolean QP (very hard problem)

$$\begin{aligned} &\min\ x^T C x \\ &\text{ s.t. } x_i \in \{-1, +1\},\ i = 1, \ldots, n \end{aligned}$$

where $C \in \mathbf{S}^n$, not necessarily PSD.

The problem can be rewritten as

$$\begin{aligned} &\min\ x^T C x \\ &\text{ s.t. } x_i^2 = 1,\ i = 1, \ldots, n \end{aligned}$$

The dual problem is shown to be an SDP

$$\begin{aligned} &\max\ -\mathbf{1}^T \nu \\ &\text{ s.t. } C + \mathbf{diag}(\nu) \succeq 0 \end{aligned}$$

where $\mathbf{1}$ is the all-one vector, & $\mathbf{diag}(\nu)$ is a diagonal matrix with diagonals given by ν.

Weak and Strong Duality

- The primal and dual optimal values

$$p^\star = \inf_{x \in C} f_0(x), \qquad d^\star = \sup_{\lambda \succeq 0, \nu} g(\lambda, \nu)$$

generally satisfy

$$d^\star \le p^\star.$$

This is called **weak duality**.

- **Strong duality** refers to cases where

$$d^\star = p^\star$$

- Strong duality does not hold for general nonconvex problems, except for some special cases.

- Strong duality **usually** holds for convex problems. (convex problems without strong duality would be pathological cases, from an application viewpoint)

Strong Duality Conditions for Convex Problems

- Constraint qualifications refer to conditions under which strong duality holds.

- **Slater's constraint qualification:** Suppose (P) is convex. If (P) is strictly feasible; i.e., there exists a point $x \in C$ such that

$$f_i(x) < 0, \quad i = 1, \ldots, m$$

  then strong duality holds.

- Slater's condition provides a very important implication that convex problems usually (though not always) have strong duality.

Examples where Strong Duality holds for a Nonconvex Problem

Minimum eigenvector problem:

$$\begin{aligned} &\min\; x^T C x \\ &\;\text{s.t.}\; x^T x = 1 \end{aligned}$$

where $C \in \mathbf{S}^n$ is not necessarily PSD. (strong duality proof is quite simple)

Nonconvex QCQP with one constraint:

$$\begin{aligned} &\min\; x^T A_0 x + 2b_0^T x + c_0 \\ &\;\text{s.t.}\; x^T A_1 x + 2b_1^T x + c_0 \leq 0 \end{aligned}$$

where $A_0, A_1 \in \mathbf{S}^n$ are not PSD. (require S-lemma to prove strong duality)

- In these special examples strong duality is quite fragile. Adding a few more constraints (even affine) could destroy strong duality.

Implication of Strong Duality in Algorithms

- Suppose that strong duality holds.
- Suppose that we can design an opt. algorithm that can produce, at iteration k,

  a primal feasible $x^{(k)}$, a dual feasible $(\lambda^{(k)}, \nu^{(k)})$

  and that it has a structure

---

repeat

$k := k + 1$.
Produce primal-dual feasible $(x^{(k)}, \lambda^{(k)}, \nu^{(k)})$ from $(x^{(k-1)}, \lambda^{(k-1)}, \nu^{(k-1)})$.

until a stopping criterion is satisfied.

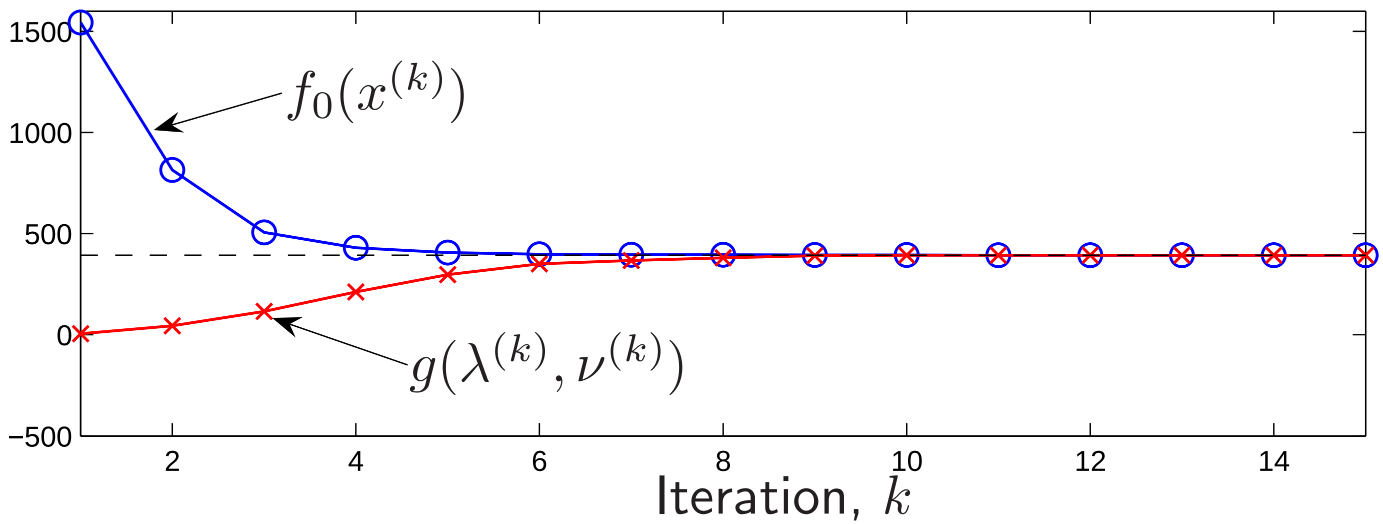
- If we stop the algorithm when

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon$$

for a given tolerance $\epsilon > 0$, then

$$f_0(x^{(k)}) - p^\star \leq \epsilon$$

that is, an ϵ-optimal solution is guaranteed.



Primal & dual values of a primal-dual path-following algorithm.

Implication of Strong Duality in Optimality Conditions

- Suppose that strong duality holds, & that $(x^\star, \lambda^\star, \nu^\star)$ is a primal-dual optimal point.

- **Complementary slackness:**

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \ldots, m$$

It implies two possibilities

$$f_i(x^\star) < 0 \implies \lambda_i^\star = 0$$
$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0$$

Solving the Primal Problem from the Dual

- Again suppose strong duality holds, & $(x^\star, \lambda^\star, \nu^\star)$ is primal-dual optimal.
- Further, suppose that $(\lambda^\star, \nu^\star)$ is known. If

$$\min_{x \in \mathcal{D}} L(x, \lambda^\star, \nu^\star) \qquad (\dagger)$$

  has a unique solution, then its solution is $x^\star$.

- Implication: If the dual problem can be easily solved, then we can solve the dual problem first, followed by solving the unconstrained minimization (†).
- Further, for a convex problem

$$L(x, \lambda^\star, \nu^\star) = f_0(x) + \textstyle\sum_i \lambda_i^\star f_i(x) + \sum_i \nu_i^\star h_i(x)$$

  is convex in x (by non-negative weighted sum). Hence (†) can be solved by

$$0 = \nabla_x L(x, \lambda^\star, \nu^\star) = \nabla f_0(x) + \textstyle\sum_i \lambda_i^\star \nabla f_i(x) + \sum_i \nu_i^\star \nabla h_i(x)$$

Karush-Kuhn-Tucker (KKT) Conditions

- Suppose f_i & h_i are differentiable.

- **For convex problems with strong duality, the sufficient & necessary conditions for $(x^\star, \lambda^\star, \nu^\star)$ to be optimal are**

$$
\begin{aligned}
h_i(x^\star) = 0, \quad f_i(x^\star) &\leq 0, \forall i && \text{(primal feasibility)} \\
\lambda_i^\star &\geq 0, \forall i && \text{(dual feasibility)} \\
\lambda_i^\star f_i(x^\star) &= 0, \forall i && \text{(complementary slackness)} \\
\nabla f_0(x^\star) + \textstyle\sum_i \lambda_i^\star \nabla f_i(x^\star) + \sum_i \nu_i^\star \nabla h_i(x^\star) &= 0
\end{aligned}
$$

- For a nonconvex problem with strong duality, KKT conditions are necessary but not sufficient.

- For a nonconvex problem with weak duality, KKT conditions are necessary conditions for local optimality with an additional assumption called *regularity*. (But if your problem only has one KKT point, then that KKT point is either optimal or there is no solution to the problem)

Importance of KKT conditions

- If the KKT conditions can be solved analytically (true only for some cases), then the problem can be solved in closed form.
- Many optimization algorithms essentially approximate the KKT conditions in an iterative fashion; e.g., the interior-point methods.

Example: Entropy maximization

$$\begin{array}{ll} \min & \sum_{i=1}^{n} x_i \log x_i \\ \text{s.t.} & \mathbf{1}^T x = 1 \end{array}$$

with domain $\mathcal{D} = \mathbf{R}^n_{++}$. The KKT equations admit a closed form

$$x_i^\star = \frac{1}{n}, \; i = 1, \dots, n$$

$$\nu^\star = \log n - 1$$

Example: Power allocation for maximum sum capacity

$$\begin{array}{ll} \max & \sum_{i=1}^{n} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) \\ \text{s.t.} & \sum_{i=1}^{n} x_i = P_{\text{total}}, \;\; x \succeq 0 \end{array}$$

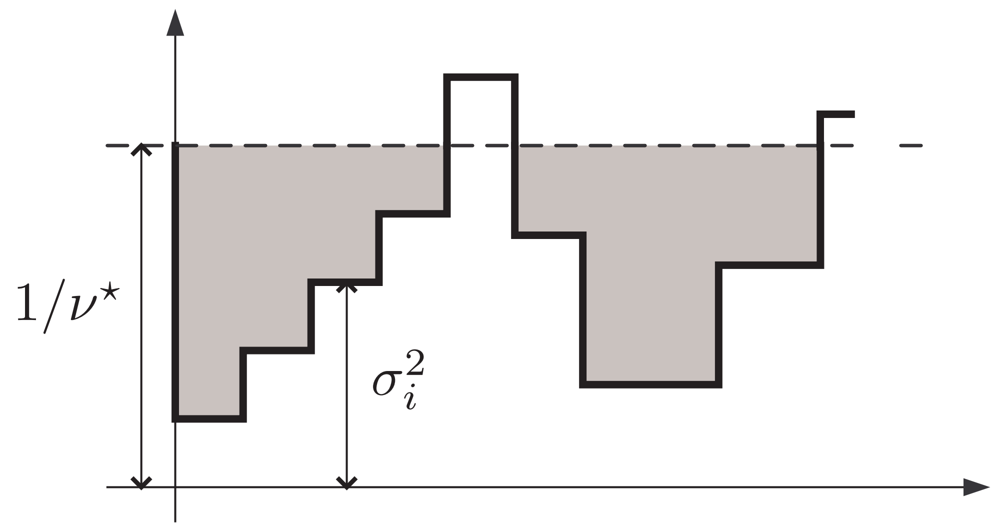
By solving the KKT conditions analytically, the solution is

$$x_i^\star = \max\left\{0, \frac{1}{\nu^\star} - \sigma_i^2\right\}$$

where $\nu^\star$ is adjusted to satisfy

$$\sum_{i=1}^{n} \max\left\{0, \frac{1}{\nu^\star} - \sigma_i^2\right\} = P_{\text{total}}$$

This is the **water-filling** solution.



Conic Optimization

- Standard conic problem

$$\begin{aligned} &\min\ c^T x \\ &\text{ s.t. } x \succeq_K 0,\ \ Ax = b \end{aligned}$$

where K is a proper cone.

- Its dual problem

$$\begin{aligned} &\max\ b^T \nu \\ &\text{ s.t. } c - A^T \nu \succeq_{K^*} 0 \end{aligned}$$

where $K^* = \{y \mid y^T x \geq 0,\ \text{for all } x \in K\}$ is the dual cone of K (convex cone).

- For $K = \mathbf{R}^n_+$ (LP), $K = \text{SOC}^n$ (SOCP), or $K = \mathbf{S}^n_+$ (SDP), K is self-dual:

$$K = K^*.$$

Example: LP. We have seen that the primal-dual pair of the standard LP is

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \succeq 0, \ Ax = b \end{array} \qquad \begin{array}{ll} \max & b^T \nu \\ \text{s.t.} & c - A^T \nu \succeq 0 \end{array}$$

Example: SDP. The standard SDP

$$\begin{array}{ll} \min & \mathbf{tr}(CX) \\ \text{s.t.} & X \succeq 0, \ \mathbf{tr}(A_i X) = b_i, \ i = 1, \ldots, m \end{array}$$

has its dual taking on the inequality form

$$\begin{array}{ll} \max & b^T \nu \\ \text{s.t.} & C - \sum_{i=1}^{m} \nu_i A_i \succeq 0 \end{array}$$

- Most properties of Lagrangian duality have their equivalent counterparts in conic opt.

- For example, Slater's condition: If there exist x such that

$$x \succ_K 0, \qquad Ax = b$$

then strong duality holds.

- KKT conditions

$$\begin{aligned} Ax^\star &= b \\ x^\star &\succeq_K 0 \\ \lambda^\star &\succeq_{K^*} 0 \\ \lambda^{\star T} x^\star &= 0 \\ c - A^T \nu^\star &= \lambda^\star \end{aligned}$$