Lagrangian Duality and Convex Programming

Lecture 12, Continuous Optimisation

Oxford University Computing Laboratory, HT 2006

Notes by Dr Raphael Hauser (hauser@comlab.ox.ac.uk)

Reformulating the KKT Conditions:

The topic of this lecture is Lagrangian duality, a generalisation of the LP duality theory we studied in the exercises relating to Lecture 8.

As a by-product of this analysis we also find that constrained convex optimisation problems allow first order necessary and sufficient conditions.

This generalises our results for unconstrained convex optimisation from Lecture 1.

We consider the constrained optimisation problem

$$
\text{(NLP)} \quad \min f(x) \quad \text{s.t.} \quad g_{\mathcal{I}}(x) \geq 0, \quad g_{\mathcal{E}}(x) = 0,
$$

where $g_{\mathcal{I}}$ is a vector of inequality constraints and $g_{\mathcal{E}}$ a vector of equality constraints.

The KKT conditions associated with this problem are

$$\nabla f(x^*) - g'_{\mathcal{I}}(x^*)^{\mathsf{T}} u^* - g'_{\mathcal{E}}(x^*)^{\mathsf{T}} v = 0, \tag{1}$$

$$g_{\mathcal{I}}(x^*) \geq 0, \tag{2}$$

$$g_{\mathcal{E}}(x^*) = 0, \tag{3}$$

$$u_j^* g_j(x^*) = 0 \qquad (j \in \mathcal{I}), \tag{4}$$

$$u^* \geq 0. \tag{5}$$

We extend the Lagrangian as follows:

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$$

$$(x, u, v) \mapsto \begin{cases} f(x) - u^{\mathsf{T}} g_{\mathcal{I}}(x) - v^{\mathsf{T}} g_{\mathcal{E}}(x), & \text{if } x \in \operatorname{dom}(f), \;\; u \geq 0, \\ +\infty & \text{if } x \notin \operatorname{dom}(f), \;\; u \geq 0, \\ -\infty & \text{if } u \ngeq 0. \end{cases}$$

This definition of the Lagrangian is a bit more general than the one we encountered previously, but this is mainly interesting for the purposes of simplifying notation and does not really entail a conceptual change.

Proposition 1: The KKT conditions (1)–(5) are equivalent to the following set of equations and inequalities,

$$\nabla_x \mathcal{L}(x^*, u^*, v^*) = 0, \tag{6}$$

$$\nabla_u \mathcal{L}(x^*, u^*, v^*) \leq 0, \tag{7}$$

$$\nabla_v \mathcal{L}(x^*, u^*, v^*) = 0, \tag{8}$$

$$u^{*\top} \nabla_u \mathcal{L}(x^*, u^*, v^*) = 0, \tag{9}$$

$$u^* \geq 0, \tag{10}$$

where $\nabla_x \mathcal{L} = (D_x \mathcal{L})^\top$ is the gradient with respect to x, and likewise $\nabla_u \mathcal{L}$ and $\nabla_v \mathcal{L}$ the gradients with respect to u and v.

Proof:

- (6) is just a reformulation of (1).

- Note that $\nabla_u \mathcal{L} = -g_{\mathcal{I}}$ and $\nabla_v \mathcal{L} = -g_{\mathcal{E}}$. Therefore, (2) is equivalent to $\nabla_u \mathcal{L}(x^*, u^*, v^*) = -g_{\mathcal{I}}(x^*) \leq 0$, which is (7).

- Likewise, (3) is equivalent to $\nabla_v \mathcal{L}(x^*, u^*, v^*) = -g_{\mathcal{E}}(x^*) = 0$, which is (8).

- Finally, (4) and $\nabla_u \mathcal{L} = -g_{\mathcal{I}}$ imply

$$u^{*\top} \nabla_u \mathcal{L}(x^*, u^*, v^*) = -\sum_{i \in \mathcal{I}} u_i^* g_i(x^*) = 0,$$

  which is (9).

- On the other hand, (7),(10) and (9) imply that $\sum_{i \in \mathcal{I}} u_i^* g_i(x^*)$ is a sum of nonnegative summands that adds to zero, and hence all the summands must be zero, which shows (4). $\square$

Proposition 2: KKT and Saddle Points.

i) Equation (6) is the first order necessary condition for x^* to be a minimiser of the unconstrained problem

$$\min_{x \in \mathbb{R}^n} \mathcal{L}(x, u^*, v^*), \tag{11}$$

where u^* and v^* are regarded as a set of fixed parameters.

ii) Equations (7)–(10) are the first order necessary optimality conditions for the problem

$$\max_{(u,v) \in \mathbb{R}^p \times \mathbb{R}^q} \mathcal{L}(x^*, u, v) \tag{12}$$

where x^* is considered as a set of fixed parameters, and where $p = |\mathcal{E}|$ and $q = |\mathcal{I}|$.

Proof:

- i) is immediate, as (11) is an unconstrained problem.

- The objective function of problem (12) takes the value $-\infty$ for $u \not\geq 0$ and finite values when $u \geq 0$. Therefore, (12) is equivalent to the constrained optimisation problem

$$\begin{aligned} \min_{(u,v)\in\mathbb{R}^p\times\mathbb{R}^q} & -\mathcal{L}(x^*, u, v) \\ \text{s.t.} \quad & u \geq 0. \end{aligned} \tag{13}$$

- The LICQ holds at all feasible points because the constraint gradients are the coordinate unit vectors $\{e_1, \dots, e_p\}$ corresponding to the variables of u, and these are linearly independent.

- The KKT conditions for (13) are therefore necessary and say that $\exists\,\lambda^* \in \mathbb{R}^p$ such that

$$\begin{bmatrix} -\nabla_u \mathcal{L}(x^*, u^*, v^*) \\ -\nabla_v \mathcal{L}(x^*, u^*, v^*) \end{bmatrix} - \sum_{j=1}^{p} \lambda_j^* e_j = 0, \tag{14}$$

$$u^* \geq 0, \qquad \text{(feasibility)} \tag{15}$$

$$\lambda_i^* u_i^* = 0 \qquad (i = 1, \ldots, p), \tag{16}$$

$$\lambda^* \geq 0. \tag{17}$$

- Equation (14) is clearly the same as

$$-\nabla_u \mathcal{L}(x^*, u^*, v^*) - \lambda^* = 0, \tag{18}$$

$$-\nabla_v \mathcal{L}(x^*, u^*, v^*) = 0, \tag{19}$$

and it is easy to see that the system (15)–(19) is equivalent to (7)–(10). □

Lagrangian Duality:

Our view of the KKT conditions in the light of Proposition 2 suggests a closer look at the saddle-point finding problems associated with $\mathcal{L}$:

$$\text{(P)} \qquad \min_{x}\left(\max_{(u,v)} \mathcal{L}(x,u,v)\right),$$

$$\text{(D)} \qquad \max_{(u,v)}\left(\min_{x} \mathcal{L}(x,u,v)\right).$$

In other words, (P) is a minimisation problem with objective function

$$x \mapsto \max_{(u,v)} \mathcal{L}(x, u, v),$$

and likewise, (D) is a maximisation problem with objective function

$$(u, v) \mapsto \min_{x} \mathcal{L}(x, u, v).$$

(P) is called the *Lagrangian primal problem* associated with (NLP) and (D) the *Lagrangian dual*.

The natural question to ask is: what is the relation between (NLP), (P) and (D)?

The following Theorem shows that (P) and (NLP) are equivalent, and later we will see that for *convex* problems (P) and (D) are equivalent under certain regularity assumptions. This amounts to showing that the max and min may be interchanged.

Theorem 1: (P) and (NLP) are equivalent problems.

Proof:

- If x is feasible for (NLP) then we have $g_{\mathcal{I}}(x) \geq 0$ and $g_{\mathcal{E}}(x) = 0$. This implies

$$\begin{aligned}\mathcal{L}(x,u,v) &= f(x) - u^{\top} g_{\mathcal{I}}(x) - v^{\top} g_{\mathcal{E}}(x) \\ &= f(x) - u^{\top} g_{\mathcal{I}}(x) \leq f(x)\end{aligned}$$

  when $u \geq 0$, and for $u \ngeq 0$ we have $\mathcal{L}(x,u,v) = -\infty$. Therefore, for (NLP)-feasible x the objective function of (P) takes the value

$$\max_{(u,v)} \mathcal{L}(x,u,v) = \mathcal{L}(x,0,v) = f(x).$$

- On the other hand, if x is infeasible for (NLP) then
  - either there exists an index $j \in \mathcal{I}$ such that $g_j(x) < 0$, and then we can choose $u_i = M > 0$,
  - or there exists an index $i \in \mathcal{E}$ such that $g_i(x) \neq 0$, and then we can choose $v_j = -\text{sign}(h_i(x))M$.

  In both cases, we can set all remaining entries of u and v to zero, and then

$$\mathcal{L}(x, u, v) \stackrel{M \to \infty}{\longrightarrow} +\infty,$$

  showing that for (NLP)-infeasible x the objective function of (P) takes the value

$$\max_{(u,v)} \mathcal{L}(x, u, v) = +\infty.$$

- In summary, we find that

$$\max_{(u,v)} \mathcal{L}(x,u,v) = \begin{cases} f(x) & \text{if } g_{\mathcal{I}}(x) \geq 0,\ g_{\mathcal{E}}(x) = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

which shows that minimising

$$x \mapsto \max_{(u,v)} \mathcal{L}(x,u,v)$$

over $\mathbb{R}^n$ is the same as minimising $f(x)$ over the feasible domain of (NLP). □

The Interpretation of the Dual:

The interpretation of the Lagrangian dual (D) is less straight forward.

Example 2.2 of the lecture notes shows that in the case where (P) is a linear programming problem, (D) is the usual LP dual, and furthermore, if (P) is a convex quadratic programming problem, then (D) is a (dual) convex quadratic programming problem.

Thus, Lagrangian duality is a generalisation of LP duality.

Weak Duality: In the LP context there was a close connection between optimality conditions and duality. We will now generalise this connection too.

Theorem 2: Weak Lagrangian Duality. For all $(x^*, u^*, v^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ it is the case that

$$\max_{(u,v)} \mathcal{L}(x^*, u, v) \geq \min_{x} \mathcal{L}(x, u^*, v^*). \tag{20}$$

Proof: This is trivial, because

$$\min_{x} \mathcal{L}(x, u^*, v^*) \leq \mathcal{L}(x^*, u^*, v^*) \leq \max_{(u,v)} \mathcal{L}(x^*, u, v). \quad \square$$

Convex Programming:

To extend the theory further, we need to assume that (NLP) is convex, that is,

- f is convex,
- g_j is concave ($j \in \mathcal{I}$),
- g_i is affine (i.e., a linear constraint) ($i \in \mathcal{E}$),

so that the feasible domain of (NLP) is convex.

Any convex programming problem is thus of the form

$$
\text{(CP)} \qquad \begin{aligned} \min_x \; & f(x) \\ \text{s.t.} \; & Ax = b, \\ & x \in \mathcal{K} = \left\{ z \in \mathbb{R}^n : \; g_j(z) \geq 0, \; (j \in \mathcal{I}) \right\}. \end{aligned}
$$

The matrix $A \in \mathbb{R}^{m \times n}$ can always be chosen so that its row vectors ∇g_i^{T} $(i \in \mathcal{E})$ are linearly independent.

$\mathcal{K}$ is a convex set.

The Lagrangian of a convex optimisation problem has nice convexity properties itself:

i) For fixed $(u^*, v^*) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ the function

$$x \mapsto \mathcal{L}(x, u^*, v^*) = f(x) + \sum_{j \in \mathcal{I}} u_j^*(-g_j(x)) + \sum_{i \in \mathcal{E}} v_i^*(-g_i(x))$$

is a sum of the convex functions f, $-u_j^* g_j$ $(j \in \mathcal{I})$ and $-v_i^* g_i$ $(i \in \mathcal{E})$. By the results of Lecture 1 this implies that

$$x \mapsto \mathcal{L}(x, u^*, v^*)$$

is globally convex!

ii) For fixed $x^* \in \mathbb{R}^n$ the function

$$(u, v) \mapsto \mathcal{L}(x^*, u, v)$$

is affine (linear plus a constant) on $\mathbb{R}^p_+ \times \mathbb{R}^q$.

Furthermore, it takes the value $-\infty$ when $u \not\geq 0$, which is consistent with our definition of concavity for so-called *proper functions* as introduced in Lecture 1. Therefore,

$$(u, v) \mapsto \mathcal{L}(x^*, u, v)$$

is globally concave!

Theorem 3: Suff. Opt. Cond. for Convex Programming. Let (NLP) be a convex problem in which the objective and constraint functions are at least once continuously differentiable. Let (x^*, u^*, v^*) be a point that satisfies the KKT conditions (6)–(10). Then x^* is a global minimiser of (NLP).

Proof:

- The condition $\nabla_x \mathcal{L}(x^*, u^*, v^*) = 0$ implies that x^* is a global minimiser of the convex unconstrained function $x \mapsto \mathcal{L}(x, u^*, v^*)$.

- For all x (NLP)-feasible we have $g_{\mathcal{I}}(x) \geq 0$ and $g_{\mathcal{E}}(x) = 0$. Since $u^* \geq 0$ we therefore have

$$\begin{aligned} f(x) &\geq f(x) - {u^*}^{\mathsf{T}} g_{\mathcal{I}}(x) - {v^*}^{\mathsf{T}} g_{\mathcal{E}}(x) \\ &= \mathcal{L}(x, u^*, v^*) \\ &\geq \mathcal{L}(x^*, u^*, v^*) \\ &= f(x^*), \end{aligned}$$

the last equality derives from the conditions (8) and (9),

$$\begin{aligned} g_{\mathcal{E}}(x^*) &= \nabla_v \mathcal{L}(x^*, u^*, v^*) = 0, \\ {u^*}^{\mathsf{T}} g_{\mathcal{I}}(x^*) &= {u^*}^{\mathsf{T}} \nabla_u \mathcal{L}(x^*, u^*, v^*) = 0. \end{aligned}$$

□

What about constraint qualifications? Why have they disappeared?!

It is important to realise that Theorem 3 only says that the KKT conditions are *sufficient* optimality conditions for convex programming, but *not necessary* conditions.

Of course, the KKT conditions also become necessary when the LICQ or the more general MFCQ is satisfied.

For convex problems it is convenient to reformulate the MFCQ by an equivalent criterion that is easier to check:

Definition 1: The convex programming problem (CP) satisfies the *Slater constraint qualification* (SCQ) if A has full row-rank and $\mathcal{K}^\circ \cap \mathcal{F}$ is nonempty, in other words, there exists a point $x \in \mathbb{R}^n$ such that $g_\mathcal{E}(x) = 0$ and $g_\mathcal{I}(x) > 0$.

Corollary 1: Convex Optimality. If (CP) satisfies the SCQ then the KKT conditions are an exact characterisation of optimality.

Proof: This follows immediately from Theorem 3 and the necessary first order optimality conditions for nonlinear programming. □

Theorem 4: Strong Lagrangian Duality. Let (CP) be a convex programming problem for which the SCQ holds and such that an optimal solution x^* exists. Then (D) has an optimal solution (u^*, v^*) and the primal and dual objective function values at x^* and (u^*, v^*) coincide.

Proof:

- Because of the SCQ, there exists a vector $(u^*, v^*) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ such that (x^*, u^*, v^*) satisfies the KKT conditions.

- Since x^* is feasible, we have

$$\begin{aligned}\mathcal{L}(x^*, u, v) &= f(x^*) - u^\top g_{\mathcal{I}}(x) - v^\top g_{\mathcal{I}}(x) \\ &= f(x^*) - u^\top g_{\mathcal{I}}(x) \\ &\leq f(x^*) \\ &= \mathcal{L}(x^*, u^*, v^*)\end{aligned}$$

  for all $(u, v) \in \mathbb{R}^p_+ \times \mathbb{R}^q$, where the last equality follows from the complementarity requirement (9) in the KKT conditions. Since $\mathcal{L}(x^*, u, v) = -\infty$ for $u \not\geq 0$, this shows that

$$\mathcal{L}(x^*, u^*, v^*) = \max_{(u,v)} \mathcal{L}(x^*, u, v).$$

- On the other hand, $\nabla_x \mathcal{L}(x^*, u^*, v^*)$ and the convexity of $x \mapsto \mathcal{L}(x, u^*, v^*)$ imply that $\mathcal{L}(x^*, u^*, v^*) = \min_x \mathcal{L}(x, u^*, v^*)$. The result now follows from weak duality. □

Reading Assignment: Lecture-Note 12.