SECTION C: CONTINUOUS OPTIMISATION
LECTURE 12: LAGRANGIAN DUALITY AND CONVEX PROGRAMMING

HONOUR SCHOOL OF MATHEMATICS, OXFORD UNIVERSITY
HILARY TERM 2005, DR RAPHAEL HAUSER

1. 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.

In all that follows we consider the constrained optimisation problem

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

where $g_{\mathcal{I}}$ is a vector of inequality constraints and $g_{\mathcal{E}}$ a vector of equality constraints. The associated KKT conditions are

$$\nabla f(x^*) - g'_{\mathcal{I}}(x^*)^{\mathrm{T}} u^* - g'_{\mathcal{E}}(x^*)^{\mathrm{T}} v = 0, \tag{1.1}$$
$$g_{\mathcal{I}}(x^*) \geq 0, \tag{1.2}$$
$$g_{\mathcal{E}}(x^*) = 0, \tag{1.3}$$
$$u_j^* g_j(x^*) = 0 \qquad (j \in \mathcal{I}), \tag{1.4}$$
$$u^* \geq 0. \tag{1.5}$$

To motivate Lagrangian duality, we will reformulate the KKT conditions (1.1)–(1.5) in slightly more abstract form. To do this, we want to extend the Lagrangian as follows:

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$$
$$(x, u, v) \mapsto \begin{cases} f(x) - u^{\mathrm{T}} g_{\mathcal{I}}(x) - v^{\mathrm{T}} g_{\mathcal{E}}(x), & \text{if } x \in \mathrm{dom}(f),\quad u \geq 0,\\ +\infty & \text{if } x \notin \mathrm{dom}(f),\quad u \geq 0,\\ -\infty & \text{if } u \ngeq 0.\end{cases}$$

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:

(i) We account for the possibility that f might not be defined on all of $\mathbb{R}^n$. Our extensions of $\mathcal{L}$ is compatible with extending f by setting $f(x) = +\infty$ for all $x \notin \mathrm{dom}\, f$. Since (NLP) is a minimisation problem, this automatically forces the search for optimal solutions to be restricted to $\mathrm{dom}\, f$.
(ii) We define $\mathcal{L}$ to be $-\infty$ when the vector of Lagrange multipliers associated with the inequality constraints is not nonnegative as it should be. Again, this convention allows us not to worry notationally about the fact that, really, u is constrained to the nonnegative orthant.

Lemma 1.1. *The KKT conditions* (1.1)–(1.5) *are equivalent to the following set of equations and inequalities,*

$$\nabla_x \mathcal{L}(x^*, u^*, v^*) = 0, \tag{1.6}$$
$$\nabla_u \mathcal{L}(x^*, u^*, v^*) \leq 0, \tag{1.7}$$
$$\nabla_v \mathcal{L}(x^*, u^*, v^*) = 0, \tag{1.8}$$
$${u^*}^{\mathrm{T}} \nabla_u \mathcal{L}(x^*, u^*, v^*) = 0, \tag{1.9}$$
$$u^* \geq 0, \tag{1.10}$$

where $\nabla_x \mathcal{L} = (D_x \mathcal{L})^{\mathrm{T}}$ 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. (1.6) is just a reformulation of (1.1). Note that $\nabla_u \mathcal{L} = -g_{\mathcal{I}}$ and $\nabla_v \mathcal{L} = -g_{\mathcal{E}}$. Therefore, (1.2) is equivalent to $\nabla_u \mathcal{L}(x^*, u^*, v^*) = -g_{\mathcal{I}}(x^*) \leq 0$, which is (1.7). Likewise, (1.3) is equivalent to $\nabla_v \mathcal{L}(x^*, u^*, v^*) = -g_{\mathcal{E}}(x^*) = 0$, which is (1.8). Finally, (1.4) and $\nabla_u \mathcal{L} = -g_{\mathcal{I}}$ imply ${u^*}^{\mathrm{T}} \nabla_u \mathcal{L}(x^*, u^*, v^*) = -\sum_{i \in \mathcal{I}} u_i^* g_i(x^*) = 0$, which is (1.9). On the other hand, (1.7),(1.10) and (1.9) 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 (1.4). □

Our reformulation of the KKT conditions in terms of the Lagrangian provides the following deeper interpretation:

Lemma 1.2 (KKT and Saddle Points).

(i) Equation (1.6) *is the first order necessary condition for x^* to be a minimiser of the unconstrained problem*

$$\min_{x \in \mathbb{R}^n} \mathcal{L}(x, u^*, v^*), \tag{1.11}$$

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

(ii) Equations (1.7)–(1.10) *are the first order necessary optimality conditions for the problem*

$$\max_{(u,v) \in \mathbb{R}^p \times \mathbb{R}^q} \mathcal{L}(x^*, u, v) \tag{1.12}$$

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

Proof. (i) (1.11) is an unconstrained problem. Therefore, (i) is immediate.

(ii) The objective function of problem (1.12) takes the value $-\infty$ for $u \ngeq 0$ and finite values when $u \geq 0$. Therefore, (1.12) is equivalent to the constrained optimisation problem

$$\begin{aligned} \min_{(u,v) \in \mathbb{R}^p \times \mathbb{R}^q} \; & -\mathcal{L}(x^*, u, v)\\ \text{s.t.}\quad & u \geq 0. \end{aligned} \tag{1.13}$$

The LICQ holds at all feasible points because the constraint gradients are the coordinate unit vectors $\{e_1, \ldots, e_p\}$ corresponding to the variables of u, and these are linearly independent. Therefore, the KKT conditions corresponding to (1.13) are necessary first order optimality conditions: the requirement is for there to exists a vector of Lagrange multipliers $\lambda^* \in \mathbb{R}^p$ such that

$$\begin{bmatrix} -\nabla_u \mathcal{L}(x^*, u^*, v^*) \\ -\nabla_v \mathcal{L}(x^*, u^*, v^*) \end{bmatrix} - \sum_{j=1}^{p} \lambda_j^* e_j = 0, \tag{1.14}$$

$$u^* \geq 0, \qquad \text{(feasibility)} \tag{1.15}$$

$$\lambda_i^* u_i^* = 0 \qquad (i = 1, \dots, p), \tag{1.16}$$

$$\lambda^* \geq 0. \tag{1.17}$$

Moreover, Equation (1.14) is clearly the same as

$$-\nabla_u \mathcal{L}(x^*, u^*, v^*) - \lambda^* = 0, \tag{1.18}$$

$$-\nabla_v \mathcal{L}(x^*, u^*, v^*) = 0, \tag{1.19}$$

Let us show that the system (1.15)–(1.19) is equivalent to (1.7)–(1.10):

- (1.18) implies $-\nabla_u \mathcal{L}(x^*, u^*, v^*) - \lambda^* = 0$ and together with (1.17) this implies $\nabla_u \mathcal{L}(x^*, u^*, v^*) \leq 0$, which is (1.7),
- (1.18) and (1.16) imply $-{u^*}^{\mathrm{T}} \nabla_u \mathcal{L}(x^*, u^*, v^*) = 0$, which implies (1.7),
- (1.19) is of course the same as (1.8), and (1.15) is the same as (1.10).

Thus, (1.15)–(1.19) imply (1.7)–(1.10). On the other hand:

- if (1.7)–(1.10) hold true then (1.19) and (1.15) hold true,
- if we set $\lambda^* = -\nabla_u \mathcal{L}(x^*, u^*, v^*)$ then (1.18) is automatically true and (1.8) implies (1.17),
- (1.9) and our choice of λ^* imply that ${\lambda^*}^{\mathrm{T}} u^* = 0$. Together with (1.10) and the already proven (1.17) this implies (1.16).

□

2. Lagrangian Duality. Our view of the KKT conditions in the light of Lemma 1.2 suggests a closer look at the saddle-point finding problems associated with $\mathcal{L}$:

$$\text{(P)} \qquad \min_x \Big(\max_{(u,v)} \ \mathcal{L}(x, u, v) \Big),$$

$$\text{(D)} \qquad \max_{(u,v)} \Big(\min_x \ \mathcal{L}(x, u, v) \Big).$$

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 regularity assumptions, that is, the max and min may be interchanged.

Theorem 2.1 (Lagrangian Primal). *(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

$$\mathcal{L}(x, u, v) = \begin{cases} f(x) - u^{\mathrm{T}} g_{\mathcal{I}}(x) - v^{\mathrm{T}} g_{\mathcal{E}}(x) = f(x) - u^{\mathrm{T}} g_{\mathcal{I}}(x) \leq f(x), & \text{if } u \geq 0, \\ -\infty & \text{if } u \ngeq 0. \end{cases}$$

Therefore, for 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 $v_j = -\operatorname{sgn}(h_i(x)) M$.

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

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

This shows that for infeasible x the objective function of (P) is

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

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). □

2.1. The Interpretation of the Dual. The interpretation of the Lagrangian dual (D) is less straight forward. The following example shows that in the case where (P) is a linear programming problem, (D) is the usual LP dual. The example also shows that convex quadratic programming problems have a convex quadratic dual. And finally, the example highlights that if (P) is not a convex problem then (D) might not yield any useful information at all.

Example 2.2. *Consider the problem*

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^{\mathrm{T}}Bx + c^{\mathrm{T}}x, \quad (2.1)$$
$$s.t. \quad Ax = b,$$
$$x \geq 0,$$

where B is a symmetric $n\times n$ matrix, $c\in\mathbb{R}^n$, A is a $q\times n$ matrix and $b\in\mathbb{R}^q$.

Problems of the form (2.1) are called *quadratic programming* (QP). We have $g_{\mathcal{I}}(x) = x$, $p = n$ and $g_{\mathcal{E}}(x) = Ax - b$. The Lagrangian of this problem is

$$\mathcal{L}(x,u,v) = \begin{cases} \frac{1}{2}x^{\mathrm{T}}Bx + (c-u-A^{\mathrm{T}}v)^{\mathrm{T}}x + b^{\mathrm{T}}v & \text{if } u\geq 0,\\ -\infty & \text{otherwise.}\end{cases}$$

Note that

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

Therefore, (P) is clearly equivalent to (2.1), as predicted by Theorem 2.1.

Let us now derive the dual of (2.1). We distinguish three cases.

Case 1: Let $B = 0$. Then (2.1) is an LP problem in standard primal form,

$$\text{(P)} \qquad \min c^{\mathrm{T}}x \quad \text{s.t. } Ax = b, \; x\geq 0.$$

In this case we have

$$\min_x \mathcal{L}(x,u,v) = \begin{cases} b^{\mathrm{T}}v & \text{if } c-u-A^{\mathrm{T}}v = 0, u\geq 0,\\ -\infty & \text{if } c-u-A^{\mathrm{T}}v \neq 0, u\geq 0,\\ -\infty & \text{if } u \ngeq 0.\end{cases}$$

or in other words,

$$\min_x \mathcal{L}(x,u,v) = \begin{cases} b^{\mathrm{T}}v & \text{if } A^{\mathrm{T}}v\leq c, u = c - A^{\mathrm{T}}v,\\ -\infty & \text{otherwise.}\end{cases}$$

Therefore, the dual Lagrangian problem is

$$\text{(D)} \qquad \max b^{\mathrm{T}}v \quad \text{s.t. } A^{\mathrm{T}}v\leq c.$$

Note that this is the usual LP dual of (P).

Case 2: Let $B\succeq 0$. If $u\geq 0$ then $x\mapsto \mathcal{L}(x,u,v)$ is a smooth convex function. The unconstrained minimisers of convex functions are exactly their stationary points characterised by $\nabla_x\mathcal{L}(x,u,v) = 0$ or

$$Bx = A^{\mathrm{T}}v + u - c, \quad (2.2)$$

so that

$$\mathcal{L}(x,u,v) = b^{\mathrm{T}}v - \frac{1}{2}x^{\mathrm{T}}Bx$$

for all such x. On the other hand, if $u\ngeq 0$ then $\min_x \mathcal{L}(x,u,v) = -\infty$. Therefore, the dual problem is

$$\max_{(x,u,v)} b^{\mathrm{T}}v - \frac{1}{2}x^{\mathrm{T}}Bx, \quad \text{s.t. } A^{\mathrm{T}}v - Bx + u = c, \; u\geq 0.$$

Moreover, if B is positive definite, then (2.2) is nonsingular, and the dual problem can be written as

$$\text{(D)} \qquad \max_{(u,v)} b^{\mathrm{T}}v - \frac{1}{2}\big(A^{\mathrm{T}}v+u-c\big)^{\mathrm{T}}B^{-1}\big(A^{\mathrm{T}}v+u-c\big), \quad \text{s.t. } u\geq 0.$$

Case 3: If B has both positive and negative eigenvalues, then $\min_x \mathcal{L}(x,u,v) = -\infty$ for all (u,v) and the dual problem becomes

$$\text{(D)} \qquad \max_{(u,v)} -\infty.$$

Showing this identity is left as an exercise. In this case (D) yields no useful information.

2.2. Weak Duality. Example 2.2 shows that Lagrangian duality is a generalisation of LP duality. In the LP context there was a close connection between optimality conditions and duality. This connection can also be generalised, as we are about to see.

Theorem 2.3 (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*

$$\max_{(u,v)} \mathcal{L}(x^*,u,v) \geq \min_x \mathcal{L}(x,u^*,v^*), \quad (2.3)$$

that is, the objective function value of (P) at the primal feasible point x^* yields an upper bound on the dual optimal value, and the objective function value of (D) at the dual feasible point (u^*,v^*) yields a lower bound on the optimal primal value.

Proof. This is trivial, because

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

□

3. Convex Programming. Weak Lagrangian duality is as far as the LP duality theory extends to nonconvex problems. To extend the theory further, we need to assume that (NLP) is convex, that is, f is convex while g_j $(j \in \mathcal{I})$ and $g_i, -g_i$ $(i \in \mathcal{E})$ are concave, so that the feasible domain $\mathcal{F}$ is convex. We have seen in Lecture 1 that the requirement that both g_i and $-g_i$ are concave implies that g_i is a linear functional plus a constant, (an affine function). Thus, *only linear equality constraints* appear in convex programming problems!

A convex programming problem is thus of the form

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

where $A \in \mathbb{R}^{m \times n}$ is a matrix which can always be chosen so that its row vectors ∇g_i^{T} $(i \in \mathcal{E})$ are linearly independent (otherwise we can eliminate a few of them or detect infeasibility), and where $\mathcal{K}$ is a convex set.

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

(i) For a 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 the convex functions f, $-u_j^* g_j$ $(j \in \mathcal{I})$ and $-v_i^* g_i$ $(i \in \mathcal{E})$. Therefore, by the results of Lecture 1, $x \mapsto \mathcal{L}(x, u^*, v^*)$ is globally convex!

(ii) For a 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 $u \ngeq 0$, and this is consistent with our definition of concavity for so-called *proper functions* as introduced in Lecture 1. Thus, $(u, v) \mapsto \mathcal{L}(x^*, u, v)$ is globally concave!

3.1. Exact Characterisation of Convex Optimality. It now turns out that – just as in unconstrained optimisation – first order optimality conditions are all we need when (NLP) is a convex problem:

Theorem 3.1 (Sufficient Optimality Conditions 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* (1.6)–(1.10). *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 feasible (for (NLP)), we have $g_{\mathcal{I}}(x) \geq 0$ and $g_{\mathcal{E}}(x) = 0$. Since $u^* \geq 0$ we therefore have

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

the last equality derives from the complementarity condition (1.9). ∎

What about constraint qualifications? Where have they disappeared to? It is important to realise that Theorem 3.1 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 3.2 (Slater Constraint Qualification). *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 3.3 (Exact Characterisation of Optimality for Convex Programming). *If (CP) satisfies the SCQ then the KKT conditions are an exact characterisation of optimality.*

Proof. This follows immediately from Theorem 3.1 and the necessary first order optimality conditions for nonlinear programming. ∎

3.2. Strong Duality for Convex Programming. In the exercises we saw that strong LP duality was a direct consequence of necessary and sufficient optimality conditions. Now that we have a generalisation of this result, strong duality extends also:

Theorem 3.4 (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

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

for all $(u, v) \in \mathbb{R}^p_+ \times \mathbb{R}^q$, where the last equality follows from the complementarity requirement (1.9) in the KKT conditions. Since $\mathcal{L}(x^*, u, v) = -\infty$ for $u \ngeq 0$, this shows that

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

On the other hand, (1.6) 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. ∎