

Lecture 7: Lagrangian duality

Michael Patriksson

2010-11-12

The Relaxation Theorem, I

- Problem: find

$$f^* := \underset{\mathbf{x}}{\text{infimum}} f(\mathbf{x}), \quad (1a)$$

$$\text{subject to } \mathbf{x} \in S, \quad (1b)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given function, $S \subseteq \mathbb{R}^n$

- A *relaxation* to (1) is a problem of the following form: to find

$$f_R^* := \underset{\mathbf{x}}{\text{infimum}} f_R(\mathbf{x}), \quad (2a)$$

$$\text{subject to } \mathbf{x} \in S_R, \quad (2b)$$

where $f_R : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function with $f_R \leq f$ on S , and $S_R \supseteq S$

The Relaxation Theorem, II

- The Relaxation Theorem:
 - (a) [relaxation] $f_R^* \leq f^*$
 - (b) [infeasibility] If (2) is infeasible, then so is (1)
 - (c) [optimal relaxation] If the problem (2) has an optimal solution, \mathbf{x}_R^* , for which it holds that

$$\mathbf{x}_R^* \in S \quad \text{and} \quad f_R(\mathbf{x}_R^*) = f(\mathbf{x}_R^*), \quad (3)$$

then \mathbf{x}_R^* is an optimal solution to (1) as well

- Proof.
- Applications: *Frank–Wolfe algorithm* (linearizing f yields lower bounds on f^* , see Chapter 12); *exterior penalty methods* yield lower bounds on f^* (see Chapter 13); *Lagrangian relaxation* yields lower bound on f^* (this chapter!)

Lagrangian relaxation, I

- Consider the optimization problem to find

$$f^* := \inf_{\mathbf{x}} f(\mathbf{x}), \quad (4a)$$

$$\text{subject to } \mathbf{x} \in X, \quad (4b)$$

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \quad (4c)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, 2, \dots, m$) are given functions, and $X \subseteq \mathbb{R}^n$

- For this problem, we assume that

$$-\infty < f^* < \infty, \quad (5)$$

that is, that f is bounded from below and that the problem has at least one feasible solution

Lagrangian relaxation, II

- For a vector $\boldsymbol{\mu} \in \mathbb{R}^m$, we define the *Lagrange function*

$$L(\mathbf{x}, \boldsymbol{\mu}) := f(\mathbf{x}) + \sum_{i=1}^m \mu_i g_i(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) \quad (6)$$

- We call the vector $\boldsymbol{\mu}^* \in \mathbb{R}_+^m$ a *Lagrange multiplier* if

$$f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*)$$

holds

Lagrange multipliers and global optima

- Let $\boldsymbol{\mu}^*$ be a Lagrange multiplier. Then, \mathbf{x}^* is an optimal solution to (4) if and only if \mathbf{x}^* is feasible in (4) and

$$\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*), \quad \text{and} \quad \mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m \quad (7)$$

- Notice the resemblance to the KKT conditions! If $X = \mathbb{R}^n$ and all functions are in C^1 then
 - the first-order optimality conditions for “ $\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*)$ ” is the same as the force equilibrium condition, the first row of the KKT conditions
 - the second item, “ $\mu_i^* g_i(\mathbf{x}^*) = 0$ for all i ” describes the complementarity conditions
- Seems to imply that there is a hidden convexity assumption here. Yes, there is: a Strong Duality Theorem is shown later

Lagrange multipliers and global optimality vs. KKT

- Extend the KKT conditions in three ways:
 - 1 $X \subseteq \mathbb{R}^n$ (KKT: $X = \mathbb{R}^n$)
 - 2 Lagrangian function $L(\cdot, \mu)$ is globally minimized (KKT: find a stationary point of the Lagrangian)
 - 3 Functions involved not necessarily in C^1
- *Note!* Works only for convex problems!
- Next: a means to find a Lagrange multiplier!

The Lagrangian dual problem, I

- Let

$$q(\boldsymbol{\mu}) := \inf_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}) \quad (8)$$

be the *Lagrangian dual function*, defined by the infimum value of the Lagrange function over X

- The *Lagrangian dual problem* is to

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad q(\boldsymbol{\mu}), \quad (9a)$$

$$\text{subject to} \quad \boldsymbol{\mu} \geq \mathbf{0}^m \quad (9b)$$

- For some $\boldsymbol{\mu}$, $q(\boldsymbol{\mu}) = -\infty$ is possible; if this is true for all $\boldsymbol{\mu} \geq \mathbf{0}^m$,

$$q^* := \sup_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu})$$

equals $-\infty$

The Lagrangian dual problem, II

- The *effective domain* of q is $D_q := \{ \mu \in \mathbb{R}^m \mid q(\mu) > -\infty \}$
- *The effective domain D_q of q is convex, and q is concave on D_q*
- *Proof.*
- That the Lagrangian dual problem always is convex (we indeed maximize a concave function!) is very good news!
- But we still need to show how a Lagrangian dual optimal solution can be used to generate a primal optimal solution

Weak Duality Theorem, I

- Let \mathbf{x} and $\boldsymbol{\mu}$ be feasible in (4) and (9), respectively. Then,

$$q(\boldsymbol{\mu}) \leq f(\mathbf{x}).$$

In particular,

$$q^* \leq f^*$$

holds

If $q(\boldsymbol{\mu}) = f(\mathbf{x})$, then the pair $(\mathbf{x}, \boldsymbol{\mu})$ is optimal in its respective problem

- *Proof.*

- The reverse direction is not always true! (Only under convexity!)

Weak Duality Theorem, II

- Weak duality is also a consequence of the Relaxation Theorem: For any $\boldsymbol{\mu} \geq \mathbf{0}^m$, let

$$S := X \cap \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \}, \quad (10a)$$

$$S_R := X, \quad (10b)$$

$$f_R := L(\boldsymbol{\mu}, \cdot) \quad (10c)$$

Apply the Relaxation Theorem

- If $q^* = f^*$, we say that there is *no (or, zero) duality gap*. If there exists a Lagrange multiplier vector, then by weak duality, this implies that there is no duality gap. The converse is not true in general: there may be cases where no Lagrange multiplier exists even when there is no duality gap; in that case though, the Lagrangian dual problem cannot have an optimal solution

On the statement of the problem (4)

- There are several ways in which the problem can be defined
- Constraints can be placed within the definition of the *ground set* X (kept intact), or within the explicit constraints defined by the functions g_i (Lagrangian relaxed)
- How to distinguish between the two, that is, how to decide whether a constraint should be kept or be Lagrangian relaxed, depends on several factors
- Keeping more constraints within X may result in a smaller duality gap, and with fewer multipliers also result in a simpler Lagrangian dual problem
- On the other hand, the Lagrangian subproblem defining the dual function then becomes more complex and difficult to solve
- There are few rules to follow: experiment and experience!

Global optimality conditions

- The following system characterizes every optimal primal and dual solution. It is applicable only in the presence of Lagrange multipliers; in other words, the system (11) is consistent if and only if there exists a Lagrange multiplier but no duality gap
- *The vector $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier if and only if*

$$\boldsymbol{\mu}^* \geq \mathbf{0}^m, \quad (\text{Dual feasibility}) \quad (11a)$$

$$\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*), \quad (\text{Lagrangian optimality}) \quad (11b)$$

$$\mathbf{x}^* \in X, \quad \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}^m, \quad (\text{Primal feasibility}) \quad (11c)$$

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m \quad (\text{Complementary slackness}) \quad (11d)$$

Saddle points

- *The vector $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier if and only if $\mathbf{x}^* \in X$, $\boldsymbol{\mu}^* \geq \mathbf{0}^m$, and $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a saddle point of the Lagrangian function on $X \times \mathbb{R}_+^m$, that is,*

$$L(\mathbf{x}^*, \boldsymbol{\mu}) \leq L(\mathbf{x}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\mu}^*), \quad (\mathbf{x}, \boldsymbol{\mu}) \in X \times \mathbb{R}_+^m, \quad (12)$$

holds

Strong duality for convex programs, introduction

- Results so far have been rather non-technical to achieve: convexity of the dual problem comes with very few assumptions on the original, primal problem, and the characterization of the primal–dual set of optimal solutions is simple and also quite easily established
- In order to establish *strong duality*, that is, to establish sufficient conditions under which there is no duality gap, however takes much more
- In particular, as is the case with the KKT conditions we need regularity conditions (that is, constraint qualifications), and we also need to utilize separation theorems

Strong duality Theorem, I

- Consider the problem (4), where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and g_i ($i = 1, \dots, m$) are convex functions and $X \subseteq \mathbb{R}^n$ is a convex set
- For this problem, we introduce the following **Slater condition**:

$$\exists \mathbf{x} \in X \text{ with } \mathbf{g}(\mathbf{x}) < \mathbf{0}^m \quad (13)$$

- Suppose that (5) and Slater's CQ (13) hold for the (convex) problem (4)
- (a) *There is no duality gap and there exists at least one Lagrange multiplier μ^* . Moreover, the set of Lagrange multipliers is bounded and convex*

Strong duality Theorem, II

- (b) *If the infimum in (4) is attained at some \mathbf{x}^* , then the pair $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ satisfies the global optimality conditions (11)*
- (c) *If the functions f and g_i are in C^1 then the condition (11b) can be written as a variational inequality. If further X is open (for example, $X = \mathbb{R}^n$) then the conditions (11) are the same as the KKT conditions*
- Similar statements for the case of also having linear equality constraints
- If all constraints are linear we can remove the Slater condition
- If f is linear then we can state the following: *If both the primal and dual problems have feasible solutions, then they both have optimal solutions, and their optimal values are the same. We will prove this elegantly in the LP chapters!*

What is the dual of an LP problem? I

- Consider the linear program to

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x}, \quad (14a)$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad (14b)$$

$$\mathbf{x} \geq \mathbf{0}^n, \quad (14c)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$. If we let $X := \mathbb{R}_+^n$, then the Lagrangian dual problem is to

$$\underset{\boldsymbol{\lambda} \in \mathbb{R}^m}{\text{maximize}} \quad \mathbf{b}^T \boldsymbol{\lambda}, \quad (15a)$$

$$\text{subject to} \quad \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c} \quad (15b)$$

- Why?

What is the dual of an LP problem? II

- The reason why we can write it in this form is that

$$q(\boldsymbol{\lambda}) := \inf_{\mathbf{x} \geq \mathbf{0}^n} \{ \mathbf{c}^T \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{b} - \mathbf{A}\mathbf{x}) \} = \mathbf{b}^T \boldsymbol{\lambda} + \inf_{\mathbf{x} \geq \mathbf{0}^n} (\mathbf{c} - \mathbf{A}^T \boldsymbol{\lambda})^T \mathbf{x},$$

so that

$$q(\boldsymbol{\lambda}) = \begin{cases} \mathbf{b}^T \boldsymbol{\lambda}, & \text{if } \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}, \\ -\infty, & \text{otherwise} \end{cases}$$

- More: $\boldsymbol{\lambda}$ unrestricted in sign, because we can write $\mathbf{A}\mathbf{x} = \mathbf{b}$ as $\mathbf{A}\mathbf{x} \leq \mathbf{b}$; $-\mathbf{A}\mathbf{x} \leq -\mathbf{b}$. Introduce $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}^-$. Lagrange function: $\mathbf{c}^T \mathbf{x} + (\boldsymbol{\mu}^+ - \boldsymbol{\mu}^-)^T (\mathbf{b} - \mathbf{A}\mathbf{x})$. Substitute $\boldsymbol{\lambda} = \boldsymbol{\mu}^+ - \boldsymbol{\mu}^-$
- If both the primal and dual problems have feasible solutions, then they both have optimal solutions, satisfying strong duality ($\mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \boldsymbol{\lambda}^*$)
- More about dual LP problems in the LP chapters!

Examples, 1: An explicit differentiable dual problem, I

- Consider the problem to

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && f(\mathbf{x}) := x_1^2 + x_2^2, \\ & \text{subject to} && x_1 + x_2 \geq 4, \\ & && x_j \geq 0, \quad j = 1, 2 \end{aligned}$$

- We consider the first constraint to be the complicated one, and hence define $g(\mathbf{x}) := -x_1 - x_2 + 4$ and let $X := \{ (x_1, x_2)^T \mid x_j \geq 0, j = 1, 2 \}$
- The Lagrangian dual function is

$$\begin{aligned} q(\mu) &= \underset{\mathbf{x} \in X}{\text{minimum}} L(\mathbf{x}, \mu) := f(\mathbf{x}) - \mu(x_1 + x_2 - 4) \\ &= 4\mu + \underset{\mathbf{x} \in X}{\text{minimum}} \{x_1^2 + x_2^2 - \mu x_1 - \mu x_2\} \\ &= 4\mu + \underset{x_1 \geq 0}{\text{minimum}} \{x_1^2 - \mu x_1\} + \underset{x_2 \geq 0}{\text{minimum}} \{x_2^2 - \mu x_2\}, \quad \mu \geq 0 \end{aligned}$$

Examples, 1: An explicit differentiable dual problem, II

- For a fixed $\mu \geq 0$, the minimum is attained at $x_1(\mu) = \frac{\mu}{2}, x_2(\mu) = \frac{\mu}{2}$
- Substituting this expression into $q(\mu)$, we obtain that $q(\mu) = f(\mathbf{x}(\mu)) - \mu(x_1(\mu) + x_2(\mu) - 4) = 4\mu - \frac{\mu^2}{2}$
- Note that q is strictly concave, and it is differentiable everywhere (due to the fact that f, g are differentiable and $\mathbf{x}(\mu)$ is unique)
- We then have that $q'(\mu) = 4 - \mu = 0 \iff \mu = 4$. As $\mu = 4 \geq 0$, it is the optimum in the dual problem!
 $\mu^* = 4; \mathbf{x}^* = (x_1(\mu^*), x_2(\mu^*))^T = (2, 2)^T$
- Also: $f(\mathbf{x}^*) = q(\mu^*) = 8$
- This is an example where the dual function is differentiable. In this particular case, the optimum \mathbf{x}^* is also unique, and is automatically given by $\mathbf{x}^* = \mathbf{x}(\mu)$
- This case is closely related to the Lagrange multiplier rule from the course on analysis in several variables

Examples, 2: An implicit non-differentiable dual problem, I

- Consider the linear programming problem to

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && f(\mathbf{x}) := -x_1 - x_2, \\ & \text{subject to} && 2x_1 + 4x_2 \leq 3, \\ & && 0 \leq x_1 \leq 2, \\ & && 0 \leq x_2 \leq 1 \end{aligned}$$

- The optimal solution is $\mathbf{x}^* = (3/2, 0)^T$, $f(\mathbf{x}^*) = -3/2$
- Consider Lagrangian relaxing the first constraint, obtaining

$$L(\mathbf{x}, \mu) = -x_1 - x_2 + \mu(2x_1 + 4x_2 - 3);$$

$$q(\mu) = -3\mu + \underset{0 \leq x_1 \leq 2}{\text{minimum}} \{(-1 + 2\mu)x_1\} + \underset{0 \leq x_2 \leq 1}{\text{minimum}} \{(-1 + 4\mu)x_2\}$$

$$= \begin{cases} -3 + 5\mu, & 0 \leq \mu \leq 1/4, \\ -2 + \mu, & 1/4 \leq \mu \leq 1/2, \\ -3\mu, & 1/2 \leq \mu \end{cases}$$

Examples, 2: An implicit non-differentiable dual problem, II

- We have that $\mu^* = 1/2$, and hence $q(\mu^*) = -3/2$. For linear programs, we have strong duality, but how do we obtain the optimal primal solution from μ^* ? q is non-differentiable at μ^* . We utilize the characterization given in (11)
- First, at μ^* , it is clear that $X(\mu^*)$ is the set $\{ \begin{pmatrix} 2\alpha \\ 0 \end{pmatrix} \mid 0 \leq \alpha \leq 1 \}$. Among the subproblem solutions, we next have to find one that is primal feasible as well as complementary
- Primal feasibility means that $2 \cdot 2\alpha + 2 \cdot 0 \leq 3 \iff \alpha \leq 3/4$
- Further, complementarity means that $\mu^* \cdot (2x_1^* + 4x_2^* - 3) = 0 \iff \alpha = 3/4$, since $\mu^* \neq 0$. We conclude that the only primal vector \mathbf{x} that satisfies the system (11) together with the dual optimal solution $\mu^* = 1/2$ is $\mathbf{x}^* = (3/2, 0)^T$
- Observe finally that $f^* = q^*$

Discussion

- As Example 2 shows the Lagrangian dual function is not always differentiable; gradient methods can then not be used
- Since it is concave, we can however utilize the theory of **supgradients** (or, **subgradients**) to develop convergent methods
- If a problem is not convex then we can have that $X(\mu^*) \cap \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \} = \emptyset$
- This means that Lagrangian relaxation cannot produce an optimal (or even a feasible) solution
- Example: integer programming. The project course is devoted partly to ways in which to obtain optimal solutions through an extended theory and/or heuristics