TMA947 / MMG621 — Nonlinear optimization

Lecture 7 — Lagrangian duality

Emil Gustavsson, Zuzana Nedělková

November 13, 2017

We will consider one the most classical versions of a general method for optimization, namely relaxation/duality based methods. The basic premise is to take a "difficult" optimization problem, and replace it with something simpler. The "simpler" here will be what we call the Lagrangian relaxation. We are now working with the problem

$$\begin{aligned} f^* = &\inf f(\boldsymbol{x}) \\ &\text{subject to } \boldsymbol{x} \in S \end{aligned} \tag{1}$$

We define a *relaxation* of the problem above to be a problem of the form

$$\begin{aligned} f_R^* = &\inf f_R(\boldsymbol{x}), \\ &\text{subject to } \boldsymbol{x} \in S_R \end{aligned} \tag{2}$$

where the function $f_R(\boldsymbol{x}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in S$, and where $S_R(\boldsymbol{x}) \subseteq S$. That is, we have replaced the feasible set with a larger one, and the objective with something smaller. The following basic result can be stated.

Theorem 1 (The relaxation theorem). *a)* $f_R^* \leq f^*$.

b) If the relaxed problem (2) *is infeasible, then so is* (1).

c) If the relaxed problem (2) *has an optimal solution* $\boldsymbol{x}_R^*$ *for which it holds that* $\boldsymbol{x}_R^* \in S$ *and* $f_R(\boldsymbol{x}_R^*) = f(\boldsymbol{x}_R^*)$*, then* $\boldsymbol{x}_R^*$ *is an optimal solution to* (1) *as well.*

Proof. See Theorem 6.1. □

1 Lagrangian relaxation

Now we consider a problem of the form

$$f^* = \inf f(\boldsymbol{x}) \tag{3a}$$
$$\text{subject to } \boldsymbol{x} \in X, \tag{3b}$$
$$g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m, \tag{3c}$$

where f and g_i are some given functions[1], and $X \subseteq \mathbb{R}^n$ is some set. The basic idea of Lagrangian relaxation is to replace *constraints* with a *price* in the objective function for their violation. That is, for any $\boldsymbol{\mu} \in \mathbb{R}^m$ we define the Lagrangian relaxation of (the constraints (3c) of) the problem (3) as the problem

$$q(\boldsymbol{\mu}) = \inf f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}), \tag{4a}$$

$$\text{subject to } \boldsymbol{x} \in X. \tag{4b}$$

Whenever $\mu_i \geq 0$ for $i = 1, \dots, m$, the above is a relaxation of (3). The objective function above is called the *Lagrange function* of (3), and is denoted as $L(\boldsymbol{x}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \sum_i \mu_i g_i(\boldsymbol{x})$.

Now we have defined a family of relaxations of (3), parametrized by the vector $\boldsymbol{\mu}$. We immediately have the following, very important, result.

Theorem 2 (Weak duality). *For any $\boldsymbol{\mu} \geq \mathbf{0}^m$, and any $\boldsymbol{x}$ feasible in* (3) *we have*

$$q(\boldsymbol{\mu}) \leq f(\boldsymbol{x}). \tag{5}$$

Proof. This is just a rephrasing of the statement that the Lagrangian relaxation is a relaxation. □

The reason why this result is so important is that it allows to get lower bounds on f^* of (3).

Example 1. *Consider the problem to*

$$f^* = \min x^2,$$
$$\text{subject to } x \geq 1.$$

Relax the constraint $x \geq 1$ to get a Lagrangian dual function (note the rewriting of $x \geq 1$ as $1 - x \leq 0$)

$$q(\mu) = \min x^2 + \mu(1-x) = \min \left(x - \frac{\mu}{2}\right)^2 - \frac{\mu^2}{4} + \mu.$$

For each fixed $\mu \geq 0$, the above is an unconstrained minimization problem of a convex function of x, so we can actually compute

$$q(\mu) = \mu - \frac{\mu^2}{4}.$$

Evaluating at, say, $\mu = 0$, we get $q(\mu) = 0$, and we can conclude that the optimal value f^* must satisfy $f^* \geq 0$. If we instead evaluate at $\mu = 1$, we would be able to conclude $f^* \geq q(1) = 3/4$.

[1] note that we do not say anything about smoothness!

Having the weak duality theorem we can try to find the *best* lower bound of f^*, which motivates the following definition.

Definition 1 (The Lagrangian dual problem)**.** *The Lagrangian dual problem to* (3) *(with respect to the relaxation of* (3c)*) is the problem*

$$q^* = \sup q(\boldsymbol{\mu}), \tag{6a}$$
$$\text{subject to } \boldsymbol{\mu} \geq \mathbf{0}^m. \tag{6b}$$

In other words, the Lagrangian dual problem is the problem of defining as tight a relaxation as possible. An immediate consequence of the weak duality theorem is: for any pair of primal/dual problems, we have $q^* \leq f^*$.

A note on terminology: from now we will refer to (6) as the dual problem, and to (3) as the primal problem. So, for example, the phrase "$\boldsymbol{x}^*$ is primally optimal" means that $\boldsymbol{x}^*$ is optimal in (3).

Example 2. *Consider again the problem from the previous example. The dual problem is to*

$$q^* = \sup_{\mu \geq 0} \mu - \frac{\mu^2}{4},$$

and one can easily verify that the maximum is attained at $\mu = 2$, $q^* = q(2) = 1$, *which can also be noted to be the optimal value* $f^* = q^* = 1$.

Note that in the above example we have $f^* = q^*$. If this holds, we say the pair of primal and dual problems has *no duality gap*. In general, we define the duality gap to be the difference $f^* - q^*$. We also define

Definition 2. *We call* $\boldsymbol{\mu}^*$ *a Lagrange multiplier vector if*

$$f^* = \inf_{x \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*). \tag{7}$$

Note that we do not have a Lagrange multiplier vector unless $f^* = q^*$. Also note the conflict of terminology, Lagrange multiplier is also used to talk about the vector $\boldsymbol{\mu}$ appearing in the Fritz-John and KKT conditions.

1.1 The dual problem

All we have done so far is to take an optimization problem (the primal) and replaced it with another problem (the dual). This only makes sense if the dual problem is in some sense "easier" than the primal.

Theorem 3. *The dual function* $q(\boldsymbol{\mu})$ *is concave, and its effective domain* $D_q = \{\boldsymbol{\mu} \mid q(\boldsymbol{\mu}) > -\infty\}$ *is convex.*

Proof. See Theorem 6.4. □

The above theorem tells us that the dual problem is a maximization of a concave function over a convex set! In other words, the dual problem is *always* a convex problem!

1.2 Global optimality conditions

If $f^* = q^*$ it turns out that we can actually use the Lagrangian relaxation to get a *sufficient* condition for optimality.

Theorem 4. *Consider the primal/dual pair of vectors $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$. Then $\boldsymbol{x}^*$ is optimal and $\boldsymbol{\mu}^*$ is a Lagrange multiplier vector if and only if*

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x} \in X}{\text{argmin}}\, L(\boldsymbol{x}, \boldsymbol{\mu}^*), \tag{8a}$$

$$\boldsymbol{\mu}^* \geq \mathbf{0}^m, \tag{8b}$$

$$\boldsymbol{x}^* \in X, \quad g_i(\boldsymbol{x}^*) \leq 0, \quad i = 1, \dots, m, \tag{8c}$$

$$\mu_i^* g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \dots, m. \tag{8d}$$

Remark: The conditions are, in order, often called Lagrangian optimality, dual feasibility, primal feasibility and complementary slackness. Note also the similarity to the KKT conditions; the only difference is that we have minimization of the Lagrangian instead of stationarity.

Proof. See Theorem 6.8. □

We can in fact formulate the above conditions in a more compact way, as what is called saddle-point optimality conditions, meaning that the pair $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ simultaneously maximizes L over $\boldsymbol{\mu}$ and minimizes L over $\boldsymbol{x}$.

Theorem 5. *$\boldsymbol{x}^*$ is primally optimal and $\boldsymbol{\mu}^*$ is a Lagrange multiplier if and only if $\boldsymbol{x}^* \in X$, $\boldsymbol{\mu}^* \geq \mathbf{0}^m$ and*

$$L(\boldsymbol{x}^*, \boldsymbol{\mu}) \leq L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \leq L(\boldsymbol{x}, \boldsymbol{\mu}^*), \quad (\boldsymbol{x}, \boldsymbol{\mu}) \in X \times \mathbb{R}^m_+. \tag{9}$$

Proof. See Theorem 6.9. □

1.3 Strong Lagrangian Duality

A natural question to ask is under what conditions one can guarantee that $q^* = f^*$, i.e., that the primal and dual optimal values coincide. It turns out that we need to require convexity and some regularity, which will here be a variant of the Slater CQ. That is, we now require that X is a convex set, g_i are convex for $i = 1 \dots, m$ and there is some point $\boldsymbol{x} \in X$ such that $g_i(\boldsymbol{x}) < 0$, $i = 1, \dots, m$.

Theorem 6. *Assume that the problem* (3) *satisfies the Slater CQ, and that $f^* \geq -\infty$. Then strong Lagrangian duality holds, and there exists at least one Lagrange multiplier vector.*

Proof. See Theorem 6.10. □

If we assume that f and g_i are also C^1, $X = \mathbb{R}^m$, and the problem (3) satisfies Slaters CQ and has some optimal solution $\boldsymbol{x}^*$. The above theorem then gives a Lagrange multiplier vector $\boldsymbol{\mu}^*$ and the pair $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of a primally optimal solution and a Lagrange multiplier vector, so it satisfy the system (8). Since the Lagrange function $L(\boldsymbol{x}, \boldsymbol{\mu})$ is convex in $\boldsymbol{x}$ for any $\boldsymbol{\mu}$, the condition that $\boldsymbol{x}^* \in \operatorname{argmin}_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*)$ can be replaced by the neccessary and sufficient condition that $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = 0$. But $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(\boldsymbol{x}^*)$. Thus in this case the global optimality conditions reduce to the Karush-Kuhn-Tucker conditions!