Lagrange duality

Geir Dahl*

February 17, 2014

This note gives a very brief introduction to Lagrange duality, and is intended for the students in the course MAT-INF3100 Linear Optimization (at the Department of Mathematics, University of Oslo). For a very good treatment of optimization theory and methods, see [1]. A much shorter presentation, with focus on convexity is [2]. Vectors in $\mathbb{R}^n$ are treated as column vectors and identified with the corresponding n-tuples.

First, from multivariate calculus, we recall the *Lagrange multiplier method.* Given functions $f, g_1, g_2, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ which are assumed to have continuous partial derivatives. Consider the optimization problem with variables $x = (x_1, x_2, \ldots, x_n)$:

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & \\ & g_1(x) = 0, \\ & \vdots \\ & g_m(x) = 0. \end{array} \tag{1}$$

Here the constraints $g_i(x) = 0$ $(i \leq m)$ are equality constraints, and a vector x satisfying all these, is called *feasible.* We look for a *local maximum*, i.e., a feasible x^* which maximizes $f(x)$ among all feasible x within a suitably small neighborhood of x^*. The main result then says that, under a weak assumption, the following holds: *If x^* is a local maximum, then there are numbers $\lambda_1, \lambda_2, \ldots, \lambda_m$ such that*

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*). \tag{2}$$

The mentioned assumption is that the gradients $\nabla g_i(x^*)$ $(i \leq m)$ are linearly independent. Here λ_i $(i \leq m)$ are called the *Lagrange multipliers.*

*University of Oslo, Dept. of Mathematics (geird@math.uio.no)

We (normally) do not know the value of the Lagrange multipliers in advance, but they are still very useful because we may solve (1) by solving a system of (possibly nonlinear) equations. This system consists of the n equations in (2) together with the m feasibility constraints $g_i(x) = 0$ $(i \leq m)$. So we have $n + m$ equations and $n + m$ unknowns (variables). If we find all solutions to this system, then among these there is a local maximum of (1). For small problems this may be done by hand, otherwise numerical methods (such as Newton's method) are used, see [1].

Next, we present how related ideas may be used to solve the more general problem where inequality constraints are present. More details may be found in [1]. So, using functions as above, consider the optimization problem

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & \\ & g_1(x) \leq 0, \\ & \vdots \\ & g_m(x) \leq 0, \\ & x \in X. \end{array} \tag{3}$$

Here $X \subseteq \mathbb{R}^n$ is any given set, representing additional constraints. For instance, $X = \mathbb{R}^n$ (no extra constraints) or $X = \{x \in \mathbb{R}^n : x \geq O\}$, or even, the more complicated case where X consists of all integral vectors in $\mathbb{R}^n$. Problem (3) is very general, and contains computationally hard problems as special cases. Although no general efficient method for solving (3) exists, the following duality approach has proved to be very useful.

Let g be the vector-valued function $g = (g_1, \ldots, g_m)$, so $g : \mathbb{R}^n \to \mathbb{R}^m$, and let $f^* = \sup\{f(x) : g(x) \leq O,\ x \in X\}$ be the optimal value in (3). Note that f^* is defined to be minus infinity if no feasible point exists, and plus infinity if there is a sequence $\{x^{(j)}\}_{j=1}^{\infty}$ of feasible points such that $f(x^{(j)}) \to \infty$. Define the *Lagrange function* $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ by

$$L(x, \lambda) = f(x) - \lambda^T g(x) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) \quad (x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m),$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$. One may think of the λ_i's as Lagrange multipliers. We also introduce the *dual function* $h : \mathbb{R}^m \to \mathbb{R}$ given by

$$h(\lambda) = \sup\{L(x, \lambda) : x \in X\} \qquad (\lambda \in \mathbb{R}^m).$$

The purpose of these concepts is to simplify the original problem and obtain upper bounds on the optimal value f^* in (3). In fact, the following result which is called *weak duality* holds:

Weak duality: *If x is feasible in* (3) *(so $x \in X$, $g(x) \leq O$), and $\lambda \geq O$, then*

$$f(x) \leq h(\lambda).$$

The proof is easy: $h(\lambda) \geq L(x, \lambda) = f(x) - \lambda^T g(x) \geq f(x)$ as $\lambda \geq O$ and $g(x) \leq O$.

If we take the supremum on the left-hand side and then the infimum on the right-hand side in the weak duality inequality, we obtain the important inequality

$$f^* \leq h^* \tag{4}$$

where $h^* = \inf\{h(\lambda) : \lambda \geq O\}$ is the optimal value of the so-called *dual problem*

$$\begin{array}{ll} \text{minimize} & h(\lambda) \\ \text{subject to} & \\ & \lambda \geq O. \end{array} \tag{5}$$

We then call (3) the *primal problem*.

As an example, consider the general linear optimization problem (see [3])

$$\max\{c^T x : Ax \leq b,\ x \geq O\}$$

where A is an $m \times n$ matrix and $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. This problem is obtained from (3) by letting $f(x) = c^T x$, $g(x) = Ax - b$ and $X = \{x \in \mathbb{R}^n : x \geq O\}$. Using the approach above we obtain

$$\begin{aligned} L(x, \lambda) &= c^T x - \lambda^T (Ax - b) \\ &= (c^T - \lambda^T A)x + \lambda^T b \\ &= (c - A^T \lambda)^T x + b^T \lambda. \end{aligned}$$

The dual function h is then

$$\begin{aligned} h(\lambda) &= \sup\{(c - A^T\lambda)^T x + b^T \lambda : x \geq O\} \\ &= \begin{cases} b^T \lambda & \text{if } A^T \lambda \geq c \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

Therefore (ignoring where $h(\lambda)$ is infinity) the dual problem becomes

$$\min\{b^T \lambda : A^T \lambda \geq c,\ \lambda \geq O\}$$

which coincides with the dual problem we know from linear optimization theory, see [3].

We conclude with some important further comments:

- *How good is the dual problem?* In other words, what can be said about the inequality $f^* \leq h^*$. If $f^* = g^*$, we say that there is no duality gap. For LP problems (as above), we have $f^* = g^*$, provided one of the problems has an optimal value. This is the *duality theorem of linear optimization*. Moreover, if one uses the simplex algorithm for solving either the primal or the dual problem, then an optimal solution for the other problem is obtained for free when the algorithm terminates (see [3]). More generally, for convex optimization, meaning that f and each g_i are convex functions (and $X = \mathbb{R}^n$), there is usually no duality gap (only a mild condition is needed for this).

- *How do we use this computationally?* Sometimes the dual function may be computed analytically, so the dual problem has an explicit form. Then, perhaps, an algorithm may be developed for the dual problem, and one obtains an upper bound h^* for the optimal value f^*. Then one may use some heuristic idea for finding a "good" primal solution x from this. More often, the dual function cannot be found analytically, and the following approach is used:

  1. Choose an initial $\lambda \geq O$.
  2. Repeat until a stopping criterion is satisfied: (a) For the given λ, solve the "relaxed subproblem": minimize $L(x, \lambda)$ over all $x \in X$, and let $x^*(\lambda)$ be an optimal solution of this subproblem. (b) Update λ using some general principle, e.g., the so-called subgradient algorithm.

  This approach, often called *Lagrangian relaxation*, has turned out to be very efficient for many large-scale problem arising in applications. There is a lot of literature on this method, both in general, and for specific applications, see [1].

References

[1] D. Bertsekas. *Nonlinear Programming*, Second Edition, Athena Scientific, 1999.

[2] G. Dahl. *An introduction to convexity.* Report 279. Dept. of Informatics, University of Oslo, 2001.

[3] R.J. Vanderbei. *Linear Programming. Foundations and Extensions*, Fouth Edition Springer, New York, 2014.