Lagrangian Duality for Dummies

David Knowles

November 13, 2010

We want to solve the following optimisation problem:

$$\min f_0(x) \tag{1}$$
$$\text{such that } f_i(x) \leq 0 \qquad \forall i \in 1, \dots, m \tag{2}$$

For now we do not need to assume convexity. For simplicity we assume no equality constraints, but all these results extend straightforwardly in that case. An obvious (but foolish) approach would be to achieve this by minimising the following function:

$$J(x) = \begin{cases} f_0(x), & \text{if } f_i(x) \leq 0 \ \forall i \\ \infty, & \text{otherwise} \end{cases} \tag{3}$$
$$= f_0(x) + \sum_i I[f_i(x)] \tag{4}$$

where $I[u]$ is a infinite step function:

$$I[u] = \begin{cases} 0, & \text{if } u \leq 0 \\ \infty, & \text{otherwise} \end{cases} \tag{5}$$

This function I gives infinite penalty to a constraint being dissatisfied. Now if we were able to minimise $J(x)$ then we would have a way of solving our optimisation problem. Unfortunately, $J(x)$ is a pretty horrible function to optimise because $I[u]$ is both non-differentiable and discontinuous. What about replacing $I[u]$ with something nicer? A straight line, λu is certainly easier to handle. This might seem like a pretty dumb choice, but for $\lambda \geq 0$ the penalty is at least in the correct direction (we are penalised for constraints being dissatisfied), and λu is a lower bound on $I[u]$ (see Figure 1). If we

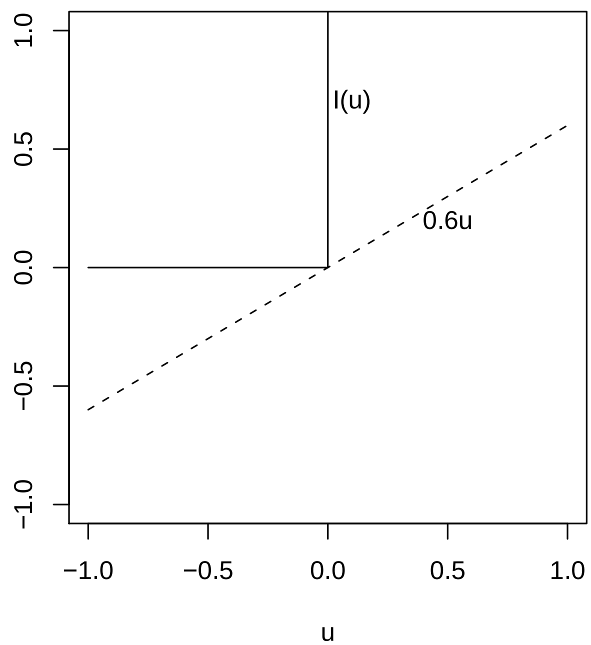

Figure 1: The infinite step function $I(u)$ and the linear relaxation λu. For $\lambda \geq 0$ note that λu is a lower bound on $I(u)$.

replace $I[u]$ by λu in the function $J(x)$ we get a function of x and λ known as the Lagrangian:

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x) \tag{6}$$

Note that if we take the maximum with respect to λ of this function we recover $J(x)$. For a particular value of x, if the constraints are all satisfied (i.e. $f_i(x) \leq 0 \ \forall i$) then the best we can do is to set $\lambda_i = 0 \ \forall i$, so $L(x, 0) = f_0(x)$. If any of the constraints are not satisfied, then $f_i(x) \geq 0$ for some i, and we can make $L(x, \lambda)$ infinite by taking $\lambda_i \to \infty$. So we have:

$$\max_\lambda L(x, \lambda) = J(x) \tag{7}$$

That's all well and good, but how does it help us solve our original optimisation problem? Remember that we want to minimise $J(x)$, which we now know means finding:

$$\min_x \max_\lambda L(x, \lambda) \tag{8}$$

This is a hard problem. But what would happen if we reversed the order of maximisation over λ and minimisation over x? Then we would be finding:

$$\max_\lambda \min_x L(x, \lambda) = \max_\lambda g(\lambda) \tag{9}$$

where $g(\lambda) = \min_x L(x, \lambda)$ is known as the dual function. Maximising the dual function $g(\lambda)$ is known as the *dual problem*, in the constrast the original *primal problem*. Since $g(\lambda)$ is a pointwise minimum of affine functions ($L(x, \lambda)$ is affine, i.e. linear, in λ), it is a concave function. The minimisation of $L(x, \lambda)$ over x might be hard. However since $g(\lambda)$ is concave and $\lambda_i \geq 0 \ \forall i$ are linear constraints, maximising $g(\lambda)$ over λ is a convex optimisation problem, i.e. an easy problem. But does solving this problem relate to the original problem? Recall that λu is a lower bound on $I(u)$. As a result, $L(x, \lambda)$ is a lower bound on $J(x)$ for all $\lambda \geq 0$. Thus

$$\begin{aligned} L(x, \lambda) &\leq J(x) & \forall \lambda \geq 0 \qquad (10)\\ \Rightarrow \min_x L(x, \lambda) &= g(\lambda) \leq \min_x J(x) =: p^\star & (11)\\ \Rightarrow d^\star &= \max_\lambda g(\lambda) \leq p^\star & (12) \end{aligned}$$

where $p^\star$ and $d^\star$ are the optima of the primal and dual problems respectively. We see that for any λ the dual function $g(\lambda)$ gives a lower bound on the optimal problem. We can then interpret the dual problem of maximising over λ as finding the tightest possible lower bound on $p^\star$:

$$\max_\lambda \min_x L(x, \lambda) \leq \min_x \max_\lambda L(x, \lambda) \tag{13}$$

This property is known as weak duality, and in fact holds in general for smooth functions L. The difference $p^\star - d^\star$ is known as the optimal duality gap. Strong duality means that we have equality, i.e. the optimal duality gap is zero. Strong duality holds if our optimisation problem is convex and a strictly feasible point exists (i.e. a point x where all constraints are strictly satisfied). In that case the solution of the primal and dual problems is equivalent, i.e. the optimal $x^\star$ is given by $\min_x L(x, \lambda^\star)$, where $\lambda^\star$ is the maximiser of $g(\lambda)$.

Duality gives us an option of trying to solve our original (potentially nonconvex) constrained optimisation problem in another way. If minimising the Lagrangian over x happens to be easy for our problem, then we know that maximising the resulting dual function over λ is easy. If strong duality holds we have found an easier approach to our original problem: if not then we still have a lower bound which may be of use. Duality also lets us formulate optimality conditions for constrained optimisation problems.

1 KKT Conditions

For an unconstrained convex optimisation problem we know we are at the global minimum if the gradient is zero. The KKT conditions are the equivalent conditions for the global minimum of a constrained convex optimisation problem. If strong duality holds and $(x^\star, \lambda^\star)$ is optimal then $x^\star$ minimises $L(x, \lambda^\star)$, giving us the first KKT condition:

$$\nabla_x L(x^\star, \lambda^\star) = \nabla_x f_0(x^\star) + \sum_i \lambda_i^\star \nabla_x f_i(x^\star) = 0 \tag{14}$$

We can interpret this condition by saying that the gradient of the objective function and constraint function must be parallel (and opposite). This means that moving along the constraint surface cannot improve the objective function. This concept is illustrated for a simple 2D optimisation problem with one inequality constraint in Figure 2.

Also, by definition we have:

$$f_0(x^\star) = g(\lambda^\star) = \min_x L(x, \lambda^\star) \leq f_0(x^\star) + \sum_i \lambda_i^\star f_i(x^\star) \leq f_0(x^\star) \tag{15}$$

where the last inequality holds because $\sum_i \lambda_i^\star f_i(x^\star) \leq 0$. We see that

$$\sum_i \lambda_i^\star f_i(x^\star) = 0 \tag{16}$$

Since $\lambda_i^\star \geq 0$ and $f_i(x^\star) \leq 0$ for all i, this gives the second KKT condition:

$$\lambda_i^\star f_i(x^\star) = 0 \qquad \forall i \tag{17}$$

This condition is known as complementary slackness: either $\lambda_i^\star$ or $f_i(x^\star)$ must be zero for all i. If the constraint is inactive then $\lambda_i^\star = 0$, and $f_i(x^\star)$ has some

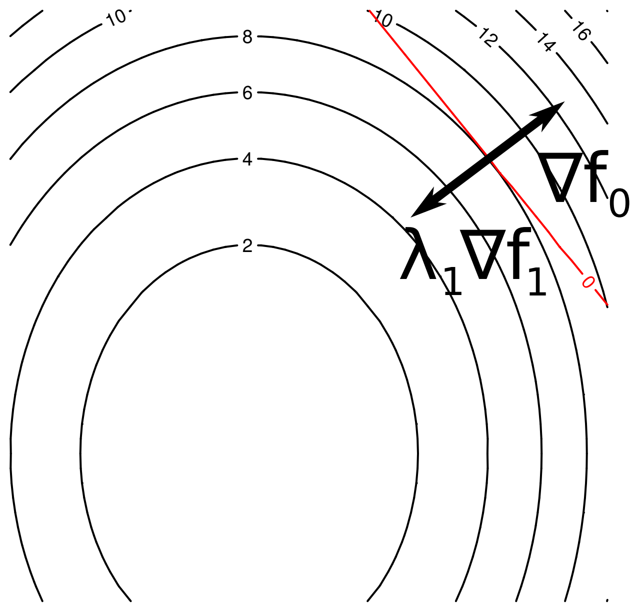

Figure 2: The first KKT condition, $\nabla_x f_0(x^\star) + \lambda_1^\star \nabla_x f_1(x^\star) = 0$. The black contours are the objective function, the red line is the constraint boundary. At the optimal solution the gradient of the objective and constraint must be parallel and opposing so that no direction along the constraint boundary could give an improved objective value.

negative value. We can interpret $\lambda_i^\star = 0$ as meaning that we can disregard constraint i in the first KKT condition since this constraint is not active. If the constraint is active then $f_i(x^\star) = 0$ and $\lambda_i^\star$ has some positive value. Note that this condition means that the lower bound λu on $I(u)$ shown in Figure 1 is tight: if λ is zero and u the functions match, and if $u = 0$, the functions also match. The remaining KKT conditions are simply the primal constraints and dual constraints (i.e. that $\lambda_i^\star \geq 0$).