7

Lagrangian Duality

Consider an optimization problem:

$$
\begin{aligned}
&\min f(x) \\
&g_i(x) \leq 0 \qquad\qquad i = 1, \dots, m \\
&h_i(x) = 0 \qquad\qquad i = 1, \dots, p
\end{aligned}
$$

with $x \in \mathbb{R}^n$. We assume its domain D to be non-empty and the problem to have a finite optimal value p^*. Note that, for the moment, we are *not* assuming convexity.

The basic idea of Lagrangian duality is to relax the problem by moving the constraints into the objective function, with some weights: the additional term in the objective should penalize the violation of said constraints. By introducing non-negative multipliers $\lambda \in \mathbb{R}^m_+$ for the m inequalities, and free multipliers $\pi \in \mathbb{R}^p$ for the p equations, we obtain the so-called Lagrangian objective function:

$$L(x, \lambda, \pi) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \pi_i h_i(x)$$

with domain $D \times \mathbb{R}^m_+ \times \mathbb{R}^p$. (λ, π) are called *Lagrange multipliers*, or *dual* variables. For fixed (λ, π), we can optimize $L(x, \lambda, \pi)$ over x, and obtain the *Lagrange dual function*:

$$l(\lambda, \pi) = \inf_{x \in D} L(x, \lambda, \pi)$$

Note that $l(\lambda, \pi)$ is defined as the pointwise infimum over all $x \in D$ of affine functions of (λ, π), and is thus a *concave* function, independent of f, g, h.

For fixed x, the terms $f(x)$, $g(x)$ and $h(x)$ are just constants in a linear expression of λ and π.

Another property of $l(\lambda, \pi)$ is that it always gives a valid lower bound on p^*.

Proposition 7.1. *For any $\lambda \geq 0$ and π, $l(\lambda, \pi) \leq p^*$.*

Proof. Let $\overline{x}$ be any feasible solution of the problem, i.e., $g(\overline{x}) \leq 0$ and $h(\overline{x}) = 0$. Then, for any $\lambda \geq 0$ and π, we have that:

$$\underbrace{\sum_{i=1}^{m} \lambda_i g_i(\overline{x})}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \pi_i h_i(\overline{x})}_{=0} \leq 0$$

Thus we have that, for any feasible $\overline{x}$:

$$l(\lambda, \pi) \leq L(\overline{x}, \lambda, \pi) \leq f(\overline{x})$$

□

In other words, the problem $\inf_{x \in D} L(x, \lambda, \pi)$ is a *relaxation* of the original problem for any $\lambda \geq 0$ and π. We call a point (λ, π) *dual-feasible* if $\lambda \geq 0$ and $l(\lambda, \pi) > -\infty$. Given that $l(\lambda, \pi)$ always provides a lower bound p^*, it makes sense to consider the problem of finding the dual multipliers associated with the best possible bound, i.e., the dual optimization problem:

Note that this is consistent with the definition of dual feasibility in the LP case.

$$\begin{gathered} \sup l(\lambda, \pi) \\ \lambda \geq 0 \end{gathered}$$

Note that this is *always* a convex optimization problem, as the domain is convex and the dual objective is concave. Let d^* be the optimal value of the Lagrange dual problem: by definition this is also a valid lower bound on p^*, and thus we have the *weak-duality* result:

$$d^* \leq p^*$$

Note that this holds even if p^* and d^* are infinite. The difference $p^* - d^*$ is called *optimal duality gap*, and it is by construction always non-negative. If the two bounds match, i.e., $p^* = d^*$, we sat that *strong duality* holds for a given problem. Strong duality does *not* hold in general, not even if we restrict to convex optimization problems. However, if all the functions involved are convex *and* we have some additional technical conditions, that go under the name of *constraint qualifications*, then we have strong duality.

7.1 Slater's conditions

We are now considering a convex optimization problem, so we have that f and g_i are all convex, and the equations $h(x) = 0$ are necessarily affine, and thus can be equivalently written as $Ax = b$.

Definition 7.1. *A point $x \in relintD$ such that $g_i(x) < 0$ and $Ax = b$ is called a* strictly feasible *solution.*

The definition could be relaxed to only require strict feasibility for nonlinear inequalities, but we will not use this in the following. The existence of a strictly feasible solution is enough to guarantee strong duality in the convex case, as stated in the following theorem:

Theorem 7.1. *If a convex optimization problem admits a strictly feasible point $\overline{x}$, then strong duality holds and it is attained, i.e., there exist (λ^*, π^*) such that $l(\lambda^*, \pi^*) = d^* = p^*$.*

Proof. To simplify the proof, let us assume that the domain D is full dimensional, so that the relative interior is actually a standard interior, and that the matrix A has full row rank p (this latter assumption can be done, as usual, without loss of generality).

Consider now the set $\mathcal{A} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$, defined as:

$$\mathcal{A} = \{(u, v, t) \mid \exists x \in D \; g(x) \leq u, Ax - b = v, f(x) \leq t\}$$

This set can be easily shown to be convex, assuming the functions f and g_i are convex themselves. Also, it has unbounded directions in the first m and in the last component, as we can always increase u and/or t while staying inside the set. Finally, we can obtain the optimal value p^* as:

Just use the definition of convex set.

$$p^* = \inf\{t \mid (0, 0, t) \in \mathcal{A}\}$$

We can also define the (simpler) set $\mathcal{B} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$ as:

$$\mathcal{B} = \{(0, 0, s) \mid s < p^*\}$$

Clearly, $\mathcal{B}$ is a convex set. In addition, $\mathcal{A}$ and $\mathcal{B}$ do not intersect: by contradiction, let $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$. Since $(u, v, t) \in \mathcal{B}$, we have $u = 0$, $v = 0$ and $t < p^*$. Since $(0, 0, t) \in \mathcal{A}$, we have that there exist $x \in D$ such that $g(x) \leq 0$, $Ax - b = 0$ and $f(x) \leq t < p^*$, which cannot be as p^* is the optimal value. So, $\mathcal{A}$ and $\mathcal{B}$ are non-empty convex set that do not intersect, so by the separating hyperplane theorem $\exists(\overline{\lambda}, \overline{\pi}, \overline{\mu}) \neq 0$ and $\alpha \in \mathbb{R}$ separating $\mathcal{A}$ from $\mathcal{B}$, i.e.:

It is just a open half line.

$$\begin{aligned} &\overline{\lambda}^\top u + \overline{\pi}^\top v + \overline{\mu} t \geq \alpha && \forall (u, v, t) \in \mathcal{A} \\ &\overline{\lambda}^\top u + \overline{\pi}^\top v + \overline{\mu} t \leq \alpha && \forall (u, v, t) \in \mathcal{B} \end{aligned}$$

The fact that $\mathcal{A}$ has unbounded directions on u and t implies that $\overline{\lambda} \geq 0$ and $\overline{\mu} \geq 0$. As for $\mathcal{B}$, the condition simplifies to $\overline{\mu} t \leq \alpha$ for any $t < p^*$, which is equivalent to $\overline{\mu} p^* \leq \alpha$. Now, for any $x \in D$, we can just pick $u = g(x)$, $v = Ax - b$ and $t = f(x)$ in $\mathcal{A}$ and obtain:

$$\overline{\lambda}^\top g(x) + \overline{\pi}^\top (Ax - b) + \overline{\mu} f(x) \geq \overline{\mu} p^* \tag{1}$$

Since $\overline{\mu} \geq 0$, there only two cases:

- $\overline{\mu} > 0$. In this case we can divide (1) by $\overline{\mu}$ and obtain:

  $$\frac{\overline{\lambda}}{\overline{\mu}}^\top g(x) + \frac{\overline{\pi}}{\overline{\mu}}^\top (Ax - b) + f(x) = L\big(x, \frac{\overline{\lambda}}{\overline{\mu}}, \frac{\overline{\pi}}{\overline{\mu}}\big) \qquad \geq p^*$$

  so we have that $L\big(x, \frac{\overline{\lambda}}{\overline{\mu}}, \frac{\overline{\pi}}{\overline{\mu}}\big) \geq p^*$ for any $x \in D$. The inequality is maintained is we minimize over $x \in D$, so we derive:

  $$l\big(\frac{\overline{\lambda}}{\overline{\mu}}, \frac{\overline{\pi}}{\overline{\mu}}\big) \geq p^*$$

  But since $l(\lambda, \pi) \leq p^*$ for any $\lambda \geq 0$ and any π, this implies that $l\big(\frac{\overline{\lambda}}{\overline{\mu}}, \frac{\overline{\pi}}{\overline{\mu}}\big) = p^*$, so strong duality holds and it is attained at $\big(\frac{\overline{\lambda}}{\overline{\mu}}, \frac{\overline{\pi}}{\overline{\mu}}\big)$.

  Notice that we haven't used the strictly feasible Slater point yet.

- $\overline{\mu} = 0$. In this case (1) simplifies to:

  $$\overline{\lambda}^\top g(x) + \overline{\pi}^\top (Ax - b) \geq 0$$

  Let us substitute the strictly feasible point $\overline{x}$ in it. We obtain $\overline{\lambda}^\top g(\overline{x}) \geq 0$, but since $g(\overline{x}) < 0$ this implies that $\overline{\lambda} = 0$. Since the separating hyperplane cannot be the zero vector, this in turns implies that $\overline{\pi} \neq 0$. In addition, we now have that $\overline{\pi}^\top (Ax - b) \geq 0$. The Slater's point satisfies $\overline{\pi}^\top (A\overline{x} - b) = 0$, and since it is in the interior of D, this means by small perturbations we either have $\overline{\pi}^\top (Ax - b) < 0$ or $\overline{\pi}^\top A = 0$: the first contradicts the inequality above, while the second contradicts the full rank assumption on A. So we conclude that $\overline{\mu} = 0$ cannot happen when the problem admits a strictly feasible point.

$\square$

7.2 Complementary slackness

In the previous section, we have found a sufficient conditions for strong duality. We now take the opposite point of view: we *assume* that strong duality holds and it is attained (both on the primal and dual side), and see which (necessary) conditions can be derived from it. Note that at the moment we are *not* assuming convexity.

Let x^* be the optimal primal solution and (λ^*, π^*) the optimal dual solution. By strong duality we have:

$$\begin{aligned} f(x^*) &= l(\lambda^*, \pi^*) \\ &= \inf_{x \in D} f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) + \sum_{i=1}^p \pi_i^* h_i(x) \\ &\leq f(x^*) + \underbrace{\sum_{i=1}^m \lambda_i^* g_i(x^*)}_{\leq 0} + \underbrace{\sum_{i=1}^p \pi_i^* h_i(x^*)}_{=0} \\ &\leq f(x^*) \end{aligned}$$

This implies that all inequalities hold at equality, hence $\sum_{i=1}^m \lambda_i^* g_i(x^*) = 0$. Since each term in the summation is non-positive, this in turn implies the m equations

$$\lambda_i^* g_i(x^*) = 0 \quad \forall i = 1, \ldots, m$$

that go under the name of *complementary slackness* conditions. The intuition is the same as in the LP case: if an inequality constraint is slack at an optimal solution, its optimal multiplier is zero, and if a dual multiplier is strictly positive, then the constraint must be tight. Another consequence of the derivation above is that x^* is a minimizer of $L(x, \lambda^*, \pi^*)$.

7.3 KKT conditions

Let us now further assume that all involved functions (f, g, h) are differentiable. Then $L(x, \lambda, \pi)$ is also differentiable w.r.t. x, and since x^* is a minimizer of $L(x, \lambda^*, \pi^*)$, then its gradient must vanish at x^*:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^p \pi_i^* \nabla h_i(x^*) = 0$$

Thus, if strong duality holds and is attained and the functions are differentiable, any optimal primal-dual pair $(x^*, (\lambda^*, \pi^*))$ *must* satisfy the following system of conditions:

$$\begin{aligned} g(x^*) &\leq 0 \\ h(x^*) &= 0 \\ \lambda^* &\geq 0 \\ \lambda_i^* g_i(x^*) &= 0 \quad i = 1, \ldots, m \\ \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^p \pi_i^* \nabla h_i(x^*) &= 0 \end{aligned} \tag{KKT}$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions. In others words, under the given assumptions, being a solution to the the KKT system is a *necessary* condition for being an optimal primal-dual pair.

If all the involved functions are convex, then we can show that any primal-dual pair satisfying the KKT system is optimal with zero duality gap. Indeed, if all the functions are convex, then we have that $L(x, \lambda^*, \pi^*)$ is itself a convex function, and its gradient vanishes at x^*, so x^* is a minimizer of $L(x, \lambda^*, \pi^*)$. In this case:

$$\begin{aligned} l(\lambda^*, \pi^*) &= L(x^*, \lambda^*, \pi^*) \\ &= f(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^* g_i(x^*)}_{=0} + \sum_{i=1}^{p} \pi_i^* \underbrace{(Ax^* - b)}_{=0} \\ &= f(x^*) \end{aligned}$$

so there is indeed no duality gap and the optimal values are attained both on the primal side, by x^*, and on the dual side, by (λ^*, π^*).

7.3.1 *Commentary on the KKT system*

The KKT system is one of the fundamental results in mathematical optimization. However, it is also one of the most misused. Let us stress here that, in general, being a solution to the KKT system does not imply anything on optimality, as the system itself is neither necessary nor sufficient for the optimality of a primal-dual pair: as a simple example, notice that if the functions are not differentiable, then the system itself cannot even be written. It is only when we make some assumptions that the KKT becomes relevant. In particular, *assuming that all the functions are differentiable*, we have:

- if strong duality holds and it is attained, then the KKT system gives a *necessary* condition.
- if the functions are convex, then the KKT system gives a *sufficient* condition.
- if the functions are convex and Slater's conditions are satisfied, then the KKT system is both necessary and sufficient.