9. Duality

Takuma Habu*
takumahabu@cornell.edu

6th December 2024

1 The Lagrange dual function

Recall the *primal* problem from static optimisation:

$$p^* := \sup_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \text{ s.t. } h_k(\mathbf{x}) = 0 \ \forall k \in \{1, 2, \dots, K\}, \qquad (1)$$
$$g_j(\mathbf{x}) \geq 0 \ \forall j \in \{1, 2, \dots, J\},$$

where $f, h_k, g_j : \mathbb{R}^d \to \mathbb{R}$ for all $k \in \{1, \dots, K\}$ and $j \in \{1, 2, \dots J\}$. We call p^* the *value of the primal*. But why do we call it the primal?

We call the problem above the primal because there is an associated problem that is called the *dual* of the primal problem. Toward defining the dual problem, recall that the Lagrangian of the primal problem is $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^K \times \mathbb{R}^J_+ \to \mathbb{R}$ given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \boldsymbol{\mu} \cdot h(\mathbf{x}) + \boldsymbol{\lambda} \cdot g(\mathbf{x}),$$

where $h = (h_k)_{k=1}^K$ and $g = (g_j)_{j=1}^J$. Define the (Lagrange) *dual function* as $q : \mathbb{R}^K \times \mathbb{R}^J_+ \to \mathbb{R}$ given by

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) := \sup_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}).$$

Observe that the dual function is the pointwise supremum of a family of affine functions of $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ and hence convex—independent of whether the primal problem is concave. Moreover, if $\mathcal{L}(\cdot, \boldsymbol{\mu}, \boldsymbol{\lambda})$ is unbounded above (in $\mathbf{x}$), then the dual function takes the value of ∞.

Why are we interested in the dual function? Let $\Gamma \subseteq \mathbb{R}^d$ denote the set of all $\mathbf{x} \in \mathbb{R}^d$ that satisfies the constraints in the primal problem. Fix any $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}^J_+$. Then, for any (feasible) $\mathbf{x} \in \Gamma$,

$$f(\mathbf{x}) \leq f(\mathbf{x}) + \boldsymbol{\mu} \cdot h(\mathbf{x}) + \boldsymbol{\lambda} \cdot g(\mathbf{x}).$$

*This note is based on "Convex Optimization" by Boyd and Vandenberghe. Like most books convex optimization, the book deals with minimisation problems. Here, we focus on maximisation problems.

Taking supremum of both sides over $\mathbf{x} \in \Gamma$ gives

$$\begin{aligned} p^* = \sup_{\mathbf{x} \in \Gamma} f(\mathbf{x}) &\leq \sup_{\mathbf{x} \in \Gamma} \; f(\mathbf{x}) + \boldsymbol{\mu} \cdot h(\mathbf{x}) + \boldsymbol{\lambda} \cdot g(\mathbf{x}) \\ &\leq \sup_{\mathbf{x} \in \mathbb{R}^d} \; f(\mathbf{x}) + \boldsymbol{\mu} \cdot h(\mathbf{x}) + \boldsymbol{\lambda} \cdot g(\mathbf{x}) = q(\boldsymbol{\mu}, \boldsymbol{\lambda}). \end{aligned}$$

We therefore conclude that

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq p^* \; \forall (\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J.$$

Whenever $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \infty$, the inequality holds vacuously. However, whenever $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < \infty$, the inequality means that the dual function gives a nontrivial upper bound of the value of the primal problem p^*.

Example 1 (LP)**.** Consider a linear programming (LP) problem:

$$\max_{\mathbf{x} \in \mathbb{R}^d} \mathbf{c}^\top \mathbf{x} \;\text{ s.t. }\; H\mathbf{x} - \mathbf{h} = \mathbf{0} \text{ and } G\mathbf{x} - \mathbf{g} \geq \mathbf{0}, \tag{2}$$

where $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{h} \in \mathbb{R}^K$, $\mathbf{g} \in \mathbb{R}^J$, $H \in \mathbb{R}^{K \times d}$ and $G \in \mathbb{R}^{J \times d}$. The Lagrangian, $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^K \times \mathbb{R}_+^J \to \mathbb{R}$ is given by

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \mathbf{c}^\top \mathbf{x} + \boldsymbol{\mu}^\top (H\mathbf{x} - \mathbf{h}) + \boldsymbol{\lambda}^\top (G\mathbf{x} - \mathbf{g}) \\ &= \left(\mathbf{c}^\top + \boldsymbol{\mu}^\top H + \boldsymbol{\lambda}^\top G\right) \mathbf{x} - \boldsymbol{\mu}^\top \mathbf{h} - \boldsymbol{\lambda}^\top \mathbf{g} \\ &= \left(\mathbf{c} + H^\top \boldsymbol{\mu} + G^\top \boldsymbol{\lambda}\right)^\top \mathbf{x} - \mathbf{h}^\top \boldsymbol{\mu} - \mathbf{g}^\top \boldsymbol{\lambda}. \end{aligned}$$

Since this is linear, the dual function is given by

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sup_{\mathbf{x} \in \mathbb{R}^n} \; \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{cases} -\mathbf{h}^\top \boldsymbol{\mu} - \mathbf{g}^\top \boldsymbol{\lambda} & \text{if } \mathbf{c} + H^\top \boldsymbol{\mu} + G^\top \boldsymbol{\lambda} = \mathbf{0} \\ \infty & \text{if } \mathbf{c} + H^\top \boldsymbol{\mu} + G^\top \boldsymbol{\lambda} \neq \mathbf{0} \end{cases}.$$

Example 2 (Least Squares)**.** Consider the problem;

$$\max_{\mathbf{x} \in \mathbb{R}^d} -\mathbf{x} \cdot \mathbf{x} \;\text{ s.t. }\; A\mathbf{x} = \mathbf{b},$$

where $A \in \mathbb{R}^{p \times d}$. This problem has p (linear) equality constraints. The Lagrangian, $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$ is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = -\mathbf{x} \cdot \mathbf{x} + \boldsymbol{\mu} \cdot (A\mathbf{x} - \mathbf{b})$$

and the dual function is given by

$$q(\boldsymbol{\mu}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \; \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}).$$

Because $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ is a concave quadratic function of $\mathbf{x}$, we can find the optimum using the first-order condition:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}) = -2\mathbf{x}^* + A^\top \boldsymbol{\mu} = \mathbf{0} \Leftrightarrow \mathbf{x}^* = \frac{1}{2} A^\top \boldsymbol{\mu}.$$

Thus, the dual function is

$$q\left(\boldsymbol{\mu}\right)=\mathcal{L}\left(\frac{1}{2}A^{\top}\boldsymbol{\mu},\boldsymbol{\mu}\right)=-\frac{1}{4}\boldsymbol{\mu}\cdot AA^{\top}\cdot\boldsymbol{\mu}-\boldsymbol{\mu}\cdot\mathbf{b}.$$

1.1 Dual function as a linear approximation of penalties

Define an indicator function $\mathbb{I}_{+},\mathbb{I}_{0}:\mathbb{R}\to\mathbb{R}$ as

$$\mathbb{I}_{+}\left(y\right)=\begin{cases}0 & \text{if } y\geq 0\\ -\infty & \text{if } y<0\end{cases},\ \mathbb{I}_{0}\left(y\right)=\begin{cases}0 & \text{if } y=0\\ -\infty & \text{if } y\neq 0\end{cases}.$$

We can use these indicator functions to rewrite the primal problem now as an unconstrained problem:

$$\max_{\mathbf{x}\in\mathbb{R}^{d}}\ f\left(\mathbf{x}\right)+\sum_{k=1}^{K}\mathbb{I}_{0}\left(h_{k}\left(\mathbf{x}\right)\right)+\sum_{j=1}^{J}\mathbb{I}_{+}\left(g_{j}\left(\mathbf{x}\right)\right).$$

The indicator functions maximally penalise violations of associated constraints—maximally because the penalty jumps from zero to minus infinity as soon as the constraints are violated, even by a little.

Now, suppose we replace $\mathbb{I}_{0}(y)$ and $\mathbb{I}_{+}(y)$ with linear functions $\mu_{k}y$ and $\lambda_{j}y$ ($\lambda_{j}>0$), respectively. Then the objective function becomes the Lagrangian function, $\mathcal{L}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda})$, and the dual function value $q(\boldsymbol{\mu},\boldsymbol{\lambda})$ is the optimal value of the problem:

$$\max_{\mathbf{x}\in\mathbb{R}^{d}}\ f\left(\mathbf{x}\right)+\sum_{k=1}^{K}\mu_{k}h_{k}\left(\mathbf{x}\right)+\sum_{j=1}^{J}\lambda_{j}g_{j}\left(\mathbf{x}\right)=\max_{\mathbf{x}\in\mathbb{R}^{d}}\ \mathcal{L}\left(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda}\right)=q\left(\boldsymbol{\mu},\boldsymbol{\lambda}\right).$$

In this formulation, we are using linear penalisations in place of indicator functions—now, the severity of the penalty grows linearly as the constraints become "more violated."

The approximation of the indicator function $\mathbb{I}_{+}(y)$ with a linear function $\lambda_{j}y$ is not great. However, the linear function is at least an over-estimator of the indicator function: because $\lambda_{j}y\geq\mathbb{I}_{+}(y)$ and $\mu_{k}y\geq\mathbb{I}_{0}(y)$ for all y, we realise that the dual function yields a lower bound on the optimal value of the original problem.

2 The (Lagrange) dual problem

We saw above that the dual function gives upper bounds on the value of the primal problem. The (*Lagrange*) *dual problem* associated with the primal problem is the problem of finding the "best" (meaning least) upper bound for the primal problem; i.e.,

$$d^{*}\coloneqq\inf_{\mathbb{R}^{K}\times\mathbb{R}_{+}^{J}}\ q\left(\boldsymbol{\mu},\boldsymbol{\lambda}\right)=\inf_{\mathbb{R}^{K}\times\mathbb{R}_{+}^{J}}\ \sup_{\mathbf{x}\in\mathbb{R}^{d}}\ \mathcal{L}\left(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda}\right)$$

We denote the value of the dual problem as d^{*}. Recall that the dual function is a supremum of a family of affine functions of $(\boldsymbol{\mu},\boldsymbol{\lambda})$ and hence convex. Thus, the dual problem is a convex problem by construction.

Example 3 (LP)**.** The Lagrange dual problem of the LP is

$$\inf_{(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathbb{R}^K\times\mathbb{R}^J} -\mathbf{h}^\top\boldsymbol{\mu} - \mathbf{g}^\top\boldsymbol{\lambda} \text{ s.t. } \mathbf{c} + H^\top\boldsymbol{\mu} + G^\top\boldsymbol{\lambda} = \mathbf{0} \text{ and } \boldsymbol{\lambda} \geq \mathbf{0}.$$

Exercise 1. What is the Lagrange dual problem of the Lagrange dual function?

2.1 Weak duality

Since every dual function is an upper bound, it follows that

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

This property, which always holds, is called *weak duality*. It holds even if d^* and p^* are infinite. For example, if the primal problem is unbounded above so that $p^* = \infty$, then $d^* = \infty$. If the dual problem is unbounded below so that $d^* = -\infty$, then $p^* = -\infty$. The latter observation, in particular, allows us to check the feasibility of p^*.

2.2 Duality gap and strong duality

The difference $d^* - p^*$ is referred to as the (optimal) *duality gap* which must be nonnegative by weakly duality. Whenever the duality gap is zero, i.e., $d^* = p^*$, we say that *strong duality* holds. Thus, whenever strong duality holds, the upper bound given by the value of the dual problem is tight. Conditions that ensure strong duality holds are called *constraint qualifications* and an example of one is the *Slater's condition* we saw when we studied KKT. To be complete, when both the objective and constraints are concave, the slater's condition holds if there exists an $\mathbf{x} \in \mathbb{R}^d$ that satisfies the constraints,[1] and

$$g_j(\mathbf{x}) > 0 \ \forall j \in \{1, 2, \ldots, J\}.$$

For any inequality constraint $j \in \{1, 2, \ldots, J\}$ for which g_j is affine, we can relax the condition above to be a weak inequality; i.e., $g_j(\mathbf{x}) \geq 0$.

Example 4 (LP)**.** Strong duality holds for any linear problem as long as the primal problem has a solution. This follows from the fact that Slater's condition reduces to feasibility (whether a point that satisfies the constraint exists) with constraints that are all affine.

Example 5 (Least Squares)**.** The Slater's condition here is simply that primal problem is feasible; i.e., $p^* = d^*$ if b is in the row space of A so that $p^* > -\infty$. In fact, for this problem, we always have strong duality even if $p^* = -\infty$; i.e., when b is not in the row space of A. The latter implies that there is a $\mathbf{z}$ with $A^\top\mathbf{z} = 0$ but $\mathbf{b}^\top\mathbf{z} \neq 0$ so that the dual function is unbounded below along the line $\{t\mathbf{z} : t \in \mathbb{R}\}$ so $d^* = -\infty$ as well.

[1] Note that while we have been assuming that the domain of objective and constraints are $\mathbb{R}^d$, in general, we need $\mathbf{x}$ to be in the relative interior (refer back to the Math Review for the definition of relative interior.) of the set of points for which the objective and all constraint functions are defined,

$$\mathcal{D} := \operatorname{dom}(f) \cap \left(\bigcap_{k=1}^{K} \operatorname{dom}(h_k)\right) \cap \left(\bigcap_{j=1}^{J} \operatorname{dom}(g_j)\right).$$

3 Minimisation problems

Given (1), let $\tilde{f} := -f$, $\tilde{h}_k := -h_k$ for all $k \in \{1, 2, \ldots, K\}$, and $\tilde{g}_j := -g_j$ for all $j \in \{1, 2, \ldots, J\}$. Suppose we consider the problem:

$$\begin{aligned}
-\tilde{p}^* = \sup_{\mathbf{x}\in\mathbb{R}^d} \tilde{f}(\mathbf{x}) \quad & \equiv - \inf_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \\
\text{s.t. } \tilde{h}_k(\mathbf{x}) = 0 \ \forall k \in \{1, 2, \ldots, K\}, \quad & \text{s.t. } h_k(\mathbf{x}) = 0 \ \forall k \in \{1, 2, \ldots, K\}, \\
\tilde{g}_j(\mathbf{x}) \geq 0 \ \forall j \in \{1, 2, \ldots, J\}. \quad & \quad g_j(\mathbf{x}) \leq 0 \ \forall j \in \{1, 2, \ldots, J\}.
\end{aligned} \tag{3}$$

Because $\sup -f = -\inf f$, above is a minimisation of f subject to the equality constraints and inequality constraints $g_j(\mathbf{x}) \leq 0$ for all $j \in \{1, 2, \ldots, J\}$ where the value of the minimised objective $\tilde{p}^*$. The Lagrangian function for the problem (3) is

$$\begin{aligned}
\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \tilde{f}(\mathbf{x}) + \boldsymbol{\mu} \cdot \tilde{h}(\mathbf{x}) + \boldsymbol{\lambda} \cdot \tilde{g}(\mathbf{x}) \\
&= -f(\mathbf{x}) - \boldsymbol{\mu} \cdot h(\mathbf{x}) - \boldsymbol{\lambda} \cdot g(\mathbf{x}) \equiv -\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}).
\end{aligned}$$

The dual function is for the problem (3)

$$\begin{aligned}
\tilde{q}(\boldsymbol{\mu}, \boldsymbol{\lambda}) &= \sup_{\mathbf{x}\in\mathbb{R}^d} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \\
&= \sup_{\mathbf{x}\in\mathbb{R}^d} -\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = -\inf_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})
\end{aligned}$$

As before, we have that

$$\begin{aligned}
\tilde{q}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = -\inf_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &\geq -\tilde{p}^* \ \forall (\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}^J_+ \\
\Leftrightarrow \inf_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &\leq \tilde{p}^* \ \forall (\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}^J_+.
\end{aligned}$$

That is, the negative of the dual function now gives us the lower bound for the value of the minimised objective $\tilde{p}^*$.

As before, we can ask what is the "best" (meaning greatest) lower bound for the value of the minimisation problem $\tilde{p}^*$:

$$\tilde{d}^* = \inf_{(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathbb{R}^K\times\mathbb{R}^J_+} \tilde{q}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sup_{(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathbb{R}^K\times\mathbb{R}^J_+} -\tilde{q}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sup_{(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathbb{R}^K\times\mathbb{R}^J_+} \inf_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}).$$

Thus, we see that the sup and inf are reversed in the case of minimisation problem.

Together, we realise that, given a minimisation problem:

$$\begin{aligned}
p_* := \inf_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \ \text{ s.t. } \ & h_k(\mathbf{x}) = 0 \ \forall k \in \{1, 2, \ldots, K\}, \\
& g_j(\mathbf{x}) \leq 0 \ \forall j \in \{1, 2, \ldots, J\}.
\end{aligned}$$

(Note the direction of the inequality constraints.) The dual function

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \coloneqq \inf_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

is a lower bound of p_* for any $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J$ and the dual problem is

$$d_* = \sup_{(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J} q(\boldsymbol{\mu}, \boldsymbol{\lambda}).$$

Weakly duality here means that $d_* \leq p_*$ and strong duality means $d_* = p_*$.

4 Interpretations of duality

4.1 Set of values

Let $\mathcal{V}$ be the set of values taken on by the constraint and objective functions:

$$\mathcal{V} = \left\{ (y, \mathbf{h}, \mathbf{g}) \in \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^J : \exists \mathbf{x} \in \mathbb{R}^d,\ (y, \mathbf{h}, \mathbf{g}) = \left(f(\mathbf{x}), \left(h_k(\mathbf{x})_{k=1}^K \right), \left(g_j(\mathbf{x})_{j=1}^J \right) \right) \right\}.$$

Then the value of the primal is given by

$$p^* = \sup \{ y : \exists \mathbf{h} = \mathbf{0} \text{ and } \mathbf{g} \geq \mathbf{0},\ (y, \mathbf{h}, \mathbf{g}) \in \mathcal{V} \}.$$

and the dual function at $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is given by

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sup_{(y, \mathbf{h}, \mathbf{g}) \in \mathcal{V}} \{ (1, \boldsymbol{\mu}, \boldsymbol{\lambda}) \cdot (y, \mathbf{h}, \mathbf{g}) \} \equiv \sup_{(y, \mathbf{h}, \mathbf{g}) \in \mathcal{V}} \left\{ y + \sum_{k=1}^K \mu_k h_k + \sum_{j=1}^J \lambda_j h_j \right\}.$$

If the dual function at $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is finite then the inequality

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq (1, \boldsymbol{\mu}, \boldsymbol{\lambda}) \cdot (y, \mathbf{h}, \mathbf{g})$$

defines hyperplane to $\mathcal{V}$ given by $H((1, \boldsymbol{\mu}, \boldsymbol{\lambda}), q(\boldsymbol{\mu}, \boldsymbol{\lambda}))$ that is supported at $(1, \boldsymbol{\mu}, \boldsymbol{\lambda})$. With $\boldsymbol{\lambda} \in \mathbb{R}_+^J$, then

$$y \leq (1, \boldsymbol{\mu}, \boldsymbol{\lambda}) \cdot (y, \mathbf{h}, \mathbf{g})$$

for any $\mathbf{g} \geq \mathbf{0}$ and $\mathbf{h} = \mathbf{0}$ so that

$$\begin{aligned} p^* &= \sup \{ y : \exists \mathbf{h} = \mathbf{0} \text{ and } \mathbf{g} \geq \mathbf{0},\ (y, \mathbf{h}, \mathbf{g}) \in \mathcal{V} \} \\ &\leq \sup \{ (1, \boldsymbol{\mu}, \boldsymbol{\lambda}) \cdot (y, \mathbf{h}, \mathbf{g}) : (y, \mathbf{h}, \mathbf{g}) \in \mathcal{V},\ \mathbf{h} = \mathbf{0} \text{ and } \mathbf{g} \geq \mathbf{0} \} \\ &\leq \sup \{ (1, \boldsymbol{\mu}, \boldsymbol{\lambda}) \cdot (y, \mathbf{h}, \mathbf{g}) : (y, \mathbf{h}, \mathbf{g}) \in \mathcal{V} \} \\ &= q(\boldsymbol{\mu}, \boldsymbol{\lambda}); \end{aligned}$$

i.e., we have weak duality.

Example 6. Consider the case in which the only constraint is a single inequality constraint (i.e., $K = 0$ and $J = 1$). Suppose we plot the value of the constraint g on the x-axis and the value of the objective y on the y-axis. The set $\mathcal{V}$ is given by some set in this space. The value of the primal problem p^* is given by the largest y such that $(y, g) \in \mathcal{V}$ for some $g \geq 0$.

Suppose we fix some $\lambda \in \mathbb{R}_+$. To obtain $q(\lambda)$, we maximise $(1, \lambda) \cdot (y, g)$ over $\mathcal{V}$ so that

$$(1, \lambda) \cdot (y, g) = y + \lambda g \leq q(\lambda) \;\; \forall (y, g) \in \mathbb{R} \times \mathbb{R}.$$

Observe that this is exactly the expression is the closed half-space below the hyperplane given by

$$H((1, \lambda), q(\lambda)) \equiv \{(y, g) \in \mathbb{R} \times \mathbb{R} : (1, \lambda) \cdot (y, g) = q(\lambda)\}.$$

This hyperplane is supported at $(y, g) \in \mathbb{R} \times \mathbb{R}$ that attains $q(\lambda)$ and has a slope $-\lambda$ and an intercept at $q(\lambda)$ because

$$(1, \lambda) \cdot (y, g) = y + \lambda g = q(\lambda) \Leftrightarrow y = -\lambda g + q(\lambda).$$

Observe that $q(\lambda)$ gives an upper bound on p^*.

4.2 Saddle point

We say that $(w^*, z^*) \in W \times Z$ is a *saddle point* of f if

$$f(w^*, z) \leq f(w, z) \leq f(w, z^*) \;\; \forall (w, z) \in W \times Z.$$

In other words, w^* minimises $f(\cdot, z')$ and z^* maximises $f(w^*, \cdot)$. Thus, when the max-min inequality holds with equality, we (alternatively) say that f has the *saddle point* property.

In Problem Set 1, we showed that, for any $f : \mathbb{R}^d \times \mathbb{R}^J \to \mathbb{R}$ with $W \subseteq \mathbb{R}^d$ and $Z \subseteq \mathbb{R}^J$,

$$\sup_{\mathbf{w} \in W} \inf_{\mathbf{z} \in Z} f(\mathbf{w}, \mathbf{z}) \leq \inf_{\mathbf{z} \in Z} \sup_{\mathbf{w} \in W} f(\mathbf{w}, \mathbf{z}).$$

This general inequality is called the *max-min inequality*. If the inequality holds with equality, we say that f has the *strong max-min property*.

Suppose we only have inequality constraints ($K = 0$). Observe that

$$\inf_{\boldsymbol{\lambda} \in \mathbb{R}^J_+} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^J_+} f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } g_j(\mathbf{x}) \geq 0 \; \forall j \in \{1, 2, \ldots, J\}, \\ -\infty & \text{otherwise.} \end{cases}$$

Hence, we can express the value of the primal problem as

$$p^* = \sup_{\mathbf{x} \in \mathbb{R}^d} \inf_{\boldsymbol{\lambda} \in \mathbb{R}^J_+} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}).$$

By definition of the dual function, we have

$$d^* = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^J_+} \sup_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}).$$

Thus, weak duality is the inequality

$$\sup_{\mathbf{x}\in\mathbb{R}^d} \inf_{\boldsymbol{\lambda}\in\mathbb{R}^J_+} \mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \leq \inf_{\boldsymbol{\lambda}\in\mathbb{R}^J_+} \sup_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \tag{4}$$

and strong duality is equivalent to the above expression holding with equality. Thus, when strong duality holds, it means that we can change the order of minimisation and maximisation without affecting the value of the problem and that we are finding $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ that solves the primal and dual problems is a saddle point of the Lagrangian.

4.3 Minimax

Consider the following two-player game that is called a *zero-sum game*. Player 1 chooses $\mathbf{w} \in W \subseteq \mathbb{R}^d$ and player 2 chooses $\mathbf{z} \in Z \subseteq \mathbb{R}^J$. Let $f(\mathbf{w}, \mathbf{z})$ denote the payoff that player 1 must make to player 2 given action profile $(\mathbf{w}, \mathbf{z})$. Then, player 1's payoff is given by f while player 2's payoff is given by $-f$. This game is called zero sum since the sum of the players' payoffs is zero. Since $\sup f = -\inf -f$, while player 1 wants to minimise f, player 2 wants to maximise f.

Suppose player 1 makes his choice first and player 2 makes his choice after observing the the choice of player 1. Player 1's best payoff in this case is

$$\sup_{\mathbf{z}\in Z} \inf_{\mathbf{w}\in W} f(\mathbf{w},\mathbf{z}).$$

If player 2 makes his choice first instead, then player 1's payoff is

$$\inf_{\mathbf{w}\in W} \sup_{\mathbf{z}\in Z} f(\mathbf{w},\mathbf{z}).$$

The max-min inequality tells us that it is always (weakly) better for a player to go second—because they are able to respond to the other play's choice. However, when f has the saddle-point property, there is no advantage to moving second. If $(\mathbf{w}^*, \mathbf{z}^*)$ is a saddle point of f, then it is called a solution of this zero-sum game.

When f is the Lagrangian function, we can think of player 1 as choosing the primal variable $\mathbf{x} \in \mathbb{R}^d$ while player 2 as choosing the dual variable $\boldsymbol{\lambda} \in \mathbb{R}^J_+$. The duality gap, $d^* - p^*$, is then equal to the advantage afforded to the player who moves second.

4.4 Price/tax

Suppose $\mathbf{x}$ represents production and $f(\mathbf{x})$ denotes the firm's profits from producing $\mathbf{x}$. Suppose each constraint $g_j(\mathbf{x}) \geq 0$ represents a limit on resources (e.g., nonnegativity constraint) or a regulatory limitation (e.g., upper bounds on how much input the firm can use). We can then interpret the following as the firm's profit maximisation problem:

$$p^* = \max_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \text{ s.t. } g_j(\mathbf{x}) \geq 0 \ \forall j \in \{1, 2, \ldots, J\}.$$

The value of the problem p^* is the firm's optimal profit.

Now suppose that the firm can pay to violate constraints by paying a cost that is linear in the size of violation as measured by g_j. Thus, the transfer made by the firm (to say the government) for violating the ith constraint is $\lambda_i f_i(\mathbf{x})$ where negative transfer represents a payment by the firm to the government and positive transfer represents the payment to the firm by the government. Assume that $\lambda_i \geq 0$ so that firm must pay for violations.

Then the Lagrangian function $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x})$ represents the profits from producing $\mathbf{x}$ net of transfers for violating limits based on constraint prices $\boldsymbol{\lambda}$. The firm wishes to maximise net profits and the dual function $q(\boldsymbol{\lambda})$ represents the optimal profit as a function of constraint prices $\boldsymbol{\lambda}$. The value of the dual problem d^* is the optimal profit under the least favourable set of constraint prices $\boldsymbol{\lambda}$.

We can think of weak duality as follows: the optimal profit when the firm can pay for constraint violations (or receive payments for non-binding constraints) is greater than or equal to the case when the firm cannot violate constraints, even with the least favourable set of constraint prices. This is because if $\mathbf{x}^*$ is optimal when firms can pay/receiver transfers for constraints, then the net profit from producing $\mathbf{x}^*$ will be lower than $f(\mathbf{x}^*)$ since some income can be derived from constraints that are not tight. The duality gap, $d^* - p^*$, is then the minimum possible advantage to the firm of being allowed to pay for constraint violations (and receive payments for nontight constraints).

Now suppose that strong duality holds and the dual optimum is attained. We can interpret dual optimal $\boldsymbol{\lambda}^*$ as the set of prices for which there is no advantage to the firm in being allowed to pay for constraint violations (or receive payments for non-binding constraints). For this reason, a dual optimal $\boldsymbol{\lambda}^*$ is sometimes called a set of shadow prices.

4.5 Certificate of suboptimality

Suppose we find a (dual) feasible $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ meaning that $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < \infty$. Then, we establish an upper bound on the value of the primal problem; i.e., $p^* \leq q(\boldsymbol{\mu}, \boldsymbol{\lambda})$. In this case, $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ provides a *certificate* that $p^* \leq q(\boldsymbol{\mu}, \boldsymbol{\lambda})$. Strong duality means that there exists arbitrarily good certificates.

Observe that feasible points allows us to bound how suboptimal a given feasible point is without knowing the exact value of p^*. To see this, take any prime feasible $\mathbf{x}$ and any dual feasible $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ and observe that

$$f(\mathbf{x}) \leq p^* \leq q(\boldsymbol{\mu}, \boldsymbol{\lambda})$$

so that

$$p^* - f(\mathbf{x}) \geq p^* - q(\boldsymbol{\mu}, \boldsymbol{\lambda}).$$

In particular, above establishes that $\mathbf{x}$ is ϵ-suboptimal with $\epsilon := p^* - q(\boldsymbol{\mu}, \boldsymbol{\lambda})$. It also establish that $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is ϵ-suboptimal for the dual problem.

If the duality gap of the primal dual feasible pair $\mathbf{x}$ and $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is zero so that $f(\mathbf{x}) = q(\boldsymbol{\mu}, \boldsymbol{\lambda})$, then $\mathbf{x}$ is primal optimal and $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is dual optimal. Thus, we can think of $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ as a certificate that proves $\mathbf{x}$ is optimal. Similarly, we can think of $\mathbf{x}$ as a certificate that proves $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ as dual optimal.

4.6 Complementary slackness

Suppose that primal and optimal values are attained and equal (which implies that strong duality holds). Let $\mathbf{x}^*$ be primal optimal and $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ be dual optimal. Then,

$$\begin{aligned}
f(\mathbf{x}^*) &= q(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \\
&= \sup_{\mathbf{x}\in\mathbb{R}^d} \left(f(\mathbf{x}) + \sum_{k=1}^{K} \mu_k^* h_k(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j^* g_j(\mathbf{x}) \right) \\
&\geq f(\mathbf{x}^*) + \sum_{k=1}^{K} \mu_k^* h_k(\mathbf{x}^*) + \sum_{j=1}^{J} \lambda_j^* g_j(\mathbf{x}^*) \\
&\geq f(\mathbf{x}^*) .
\end{aligned}$$

The first line is the statement that the duality gap is zero and the second line follows from the definition of the dual function. The third line follows because the supremum of the Lagrangian over $\mathbf{x}$ is les than its value at $\mathbf{x} = \mathbf{x}^*$. The last inequality follows from $\lambda_j^*, g_j(\mathbf{x}^*) \geq 0$ for all $j \in \{1, 2 \ldots, J\}$ and $h_k(\mathbf{x}^*) = 0$ for all $k \in \{1, 2, \ldots, K\}$. Of course the chain of equality means that the inequalities are in fact equalities. Note:

- ▷ the third inequality (which is in fact an equality) tells us that $\mathbf{x}^*$ maximises $\mathcal{L}(\cdot, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$;
- ▷ The last equality gives us complementary slackness:

$$\sum_{j=1}^{J} \lambda_j^* g_j(\mathbf{x}^*) = 0 \Leftrightarrow \lambda_j^* g_j(\mathbf{x}^*) = 0 \ \forall j \in \{1, 2, \ldots, J\}$$

  because each λ_j and $g_j(\mathbf{x}^*)$ are nonnegative.

4.7 KKT conditions

Suppose f, h and g are differentiable. Let $\mathbf{x}^*$ and $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ be any primal and dual optimal points with zero duality gap. Since $\mathbf{x}^*$ maximises $\mathcal{L}(\cdot, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$, necessary condition for optimum means

$$0 = \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \nabla f(\mathbf{x}^*) + \sum_{k=1}^{K} \mu_k^* \nabla h_k(\mathbf{x}^*) + \sum_{j=1}^{J} \lambda_j^* \nabla g_j(\mathbf{x}^*) .$$

Recall that KKT conditions consists of the above condition and

(i) primal feasibility: $h_k(\mathbf{x}^*) = 0$ for all $k \in \{1, 2, \ldots, K\}$ and $g_j(\mathbf{x}^*) \geq 0$ for all $j \in \{1, 2, \ldots, J\}$;

(ii) dual feasibility: $\lambda_j^* \geq 0$ for all $j \in \{1, 2, \ldots, J\}$;

(iii) complementary slackness: $\lambda_j^* g_j^*(\mathbf{x}^*) = 0$ for all $j \in \{1, 2, \ldots, J\}$.

Notice that (i) ensures that $\mathbf{x}^*$ is primal feasible. Together with the first-order condition, we have

$$\begin{aligned}
q(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) &= \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \\
&= f(\mathbf{x}^*) + \sum_{k=1}^{K} \mu_k^* h_k(\mathbf{x}^*) + \sum_{j=1}^{J} \lambda_j^* g_j(\mathbf{x}^*) = f(\mathbf{x}^*) = p^*.
\end{aligned}$$

This tells us that duality gap is indeed zero.

5 Theorems of alternatives

5.1 Weak alternatives

Two systems of inequalities are called *weak alternatives* if at most one of two is feasible.

Consider the primal problem with $f(\cdot) \coloneqq 0$; i.e.,

$$\begin{aligned} \sup_{\mathbf{x} \in \mathbb{R}^d} 0 \;\; \text{s.t.} \;\; & h_k(\mathbf{x}) = 0 \;\forall k \in \{1, 2, \ldots, K\}, \\ & g_j(\mathbf{x}) \geq 0 \;\forall j \in \{1, 2, \ldots, J\}, \end{aligned} \tag{5}$$

Let

$$\Gamma = \left\{\mathbf{x} \in \mathbb{R}^d : h_k(\mathbf{x}) = 0 \;\forall k \in \{1, \ldots, K\},\; g_j(\mathbf{x}) \geq 0 \;\forall j \in \{1, \ldots, J\}\right\}$$

denote the set of primal variables that satisfy the constraint. The value of this problem is

$$p^* = \begin{cases} 0 & \text{if } \Gamma \neq \varnothing \\ -\infty & \text{if } \Gamma = \varnothing \end{cases}.$$

Thus, we can determine whether a system of constraints is feasible, i.e., $\Gamma \neq \varnothing$, by solving the optimisation problem above.

In this case, the dual function is

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \sum_{k=1}^{K} \mu_k h_k(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x})$$

and the dual problem is to maximise q with respect to $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J$. Observe that q is homogenous, i.e., $q(\alpha\boldsymbol{\mu}, \alpha\boldsymbol{\lambda}) = \alpha q(\boldsymbol{\mu}, \boldsymbol{\lambda})$, so that

$$d^* = \begin{cases} -\infty & \text{if } \exists\,(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J,\; q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Weak duality tells us that $p^* \leq d^*$. Hence, if the we can find $(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}_+^J$ such that $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < 0$, then the primal system of constraints must be infeasible; i.e., $\Gamma = \varnothing$. That is, finding such a feasible $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is a certificate of infeasibility of Γ.

Notice also that if $\Gamma \neq \varnothing$ so that the primal system of constraints is feasible. then the constraints

$$\Lambda \coloneqq \left\{(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}^J :\; \boldsymbol{\lambda} \geq \mathbf{0},\; q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < 0\right\}$$

must be infeasible. So we can interpret existence of an $\mathbf{x} \in \Gamma$ as a certificate of infeasibility of the system of constraints in the dual problem. Thus, we conclude that the sets of inequalities that define Γ and Λ are weak alternatives—this is the case whether the g_j's are concave or h_k's are affine.

5.2 Strong alternatives

Suppose now that g_j's are concave and h_k's are affine. If a constraint qualification holds so that strong duality holds, then pairs of weak alternatives are called *strong alternatives*. It means that *exactly one* of the two alternatives holds—e.g., a system of inequalities is feasible if and only if the other is infeasible.

Let us consider the following system of inequalities:

$$g_j(\mathbf{x}) \geq 0 \ \forall j \in \{1, 2, \ldots, J\} \ \text{ and } A\mathbf{x} = \mathbf{b}, \tag{6}$$

where each g_j is concave, and its alternative

$$\boldsymbol{\lambda} \geq \mathbf{0} \text{ and } q(\boldsymbol{\mu}, \boldsymbol{\lambda}) < 0. \tag{7}$$

Let us show that the two sets of inequalities are strong alternatives under certain assumptions.

Consider the following problem:

$$p^* = \sup_{s \in \mathbb{R},\, \mathbf{x} \in \mathbb{R}^d} s \ \text{ s.t. } \ A\mathbf{x} = \mathbf{b}, \ g_j(\mathbf{x}) \geq s \ \forall j \in \{1, 2, \ldots, J\}. \tag{8}$$

The value of this problem p^* is strictly positive if and only if there exists a solution to the strict inequality system (4). We will assume that there exists $\mathbf{x} \in \mathbb{R}^d$ that satisfies $A\mathbf{x} = \mathbf{b}$,[2] and that p^* is attained by some $(s, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^d$.[3]

The Lagrange dual function of this problem is

$$\sup_{\mathbf{x} \in \mathbb{R}^d,\ s \in \mathbb{R}} s + \sum_{j=1}^{J} \lambda_j (g_j(\mathbf{x}) - s) + \boldsymbol{\mu}^\top (A\mathbf{x} - \mathbf{b}) = \begin{cases} q(\boldsymbol{\mu}, \boldsymbol{\lambda}) & \sum_{j=1}^{J} \lambda_j = 1, \\ \infty & \text{otherwise.} \end{cases}$$

Observe that $\sum_{j=1}^{J} \lambda_j = 1$ ensures that s cancels out in the objective. We can express the dual problem as

$$\inf_{(\boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbb{R}^K \times \mathbb{R}^J} q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \ \text{ s.t. } \ \boldsymbol{\lambda} \geq 0 \text{ and } \sum_{j=1}^{J} \lambda_j = 1.$$

Observe that the Slater's condition holds for (8): by hypothesis, we can find $\mathbf{x} \in \mathbb{R}^d$ such that $A\mathbf{x} = \mathbf{b}$ and choosing any $s < \min_j g_j(\mathbf{x})$ yields a point that is strictly feasible. Therefore, we have $d^* = p^*$ and the dual optimum d^* is attained; i.e., there exists $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ such that

$$q(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = p^*, \ \boldsymbol{\lambda}^* \geq 0 \text{ and } \sum_{j=1}^{J} \lambda_j^* = 1.$$

Now suppose that system of inequalities (6) is infeasible so that $p^* < 0$ (her we are using the fact that p^* is attained). Then, $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ satisfy the alternative inequality system (7). Alternatively, if (7) is feasible, then $d^* = p^* < 0$ which shows that the (6) is infeasible. Thus, the two system of inequalities (6) and (7) are strong alternatives.

[2]When the domain of objective and constraints are not necessarily $\mathbb{R}^d$, then we need to also impose that $\mathbf{x}$ that solves $A\mathbf{x} = \mathbf{b}$ also is in the relative interior of $\mathcal{D}$.

[3]For example, p^* is attained if $\min_j g_j(\mathbf{x}) \to -\infty$ as $\mathbf{x} \to \infty$.

5.3 Farkas' lemma

Theorem 1 (Farkas' lemma)**.** *The following two systems of inequalities are strong alternatives:*

(i) $A\mathbf{x} \geq \mathbf{0}$ *and* $\mathbf{c}^\top \mathbf{x} > 0$*, where* $A \in \mathbb{R}^{J \times d}$ *and* $\mathbf{c} \in \mathbb{R}^d$*; and*

(ii) $\mathbf{c} + A^\top \boldsymbol{\lambda} = \mathbf{0}$ *and* $\boldsymbol{\lambda} \geq 0$*.*

Proof. Consider the following primal LP:

$$p^* = \max_{\mathbf{x} \in \mathbb{R}^d} \mathbf{c}^\top \mathbf{x} \ \text{ s.t. } \ A\mathbf{x} \geq \mathbf{0}.$$

The dual problem is given by

$$d^* = \inf_{\boldsymbol{\lambda}} 0 \text{ s.t. } \mathbf{c} + A^\top \boldsymbol{\lambda} = \mathbf{0} \text{ and } \boldsymbol{\lambda} \geq \mathbf{0}.$$

Because $\mathbf{x} = \mathbf{0}$ is feasible in the primal LP, we can rule out the one case in which strong duality can fail for LPs. Hence, we must have $p^* = d^*$. Now observe that

- ▷ The primal LP has optimal value 0 if (i) is not feasible and optimal value ∞ if (i) is feasible.
- ▷ The dual problem has optimal value 0 if (ii) is feasible and value ∞ if (ii) is infeasible.

Hence, if (i) is feasible so that $p^* = \infty = d^*$, then (ii) is infeasible. Alternatively, if (i) is infeasible so that $p^* = 0 = d^*$, then (ii) is feasible. Together, these imply that (i) and (ii) are strong alternatives. ■

6 A proof of strong duality

Consider a primal problem in which f and g are concave and h is affine. We write the affine equality constraint as $A\mathbf{x} = \mathbf{b}$. Suppose further that the following Slater's condition holds: there exists $\mathbf{x} \in \mathbb{R}^d$ with $g_j(\mathbf{x}) > 0$ for all $j \in \{1, 2, \ldots, J\}$. To simplify the proof, assume $\text{rank}(A) = K$. Suppose p^* is finite (note that $p^* > -\infty$ because there is a feasible point by Slater's condition. While it's possible that $p^* = \infty$, weakly duality implies $d^* = \infty$ so that $d^* = p^*$ in this case).

Recall that $\mathcal{V}$ is the set of values taken on by the constraint and objective functions:

$$\mathcal{V} = \left\{ (y, \mathbf{h}, \mathbf{g}) \in \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^J : \exists \mathbf{x} \in \mathbb{R}^d, \ (y, \mathbf{h}, \mathbf{g}) = \left(f(\mathbf{x}), \left(h_k(\mathbf{x})_{k=1}^K \right), \left(g_j(\mathbf{x})_{j=1}^J \right) \right) \right\}.$$

Let

$$\mathcal{A} := \{ (y, \mathbf{h}, \mathbf{g}) \in \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^J : \exists \mathbf{x} \in \mathbb{R}^d, \ f(\mathbf{x}) \geq y, \ h_k(\mathbf{x}) = h_k \ \forall k \in \{1, 2 \ldots, K\}, \\ g_j(\mathbf{x}) \geq g_j \ \forall j \in \{1, 2, \ldots, J\} \}.$$

That is, $\mathcal{A}$ contains all the points in $\mathcal{V}$ as well as points that are "worse", i.e., those with smaller objective or inequality constraint function values. Clearly, $\mathcal{A}$ is convex when f and g are concave and h is affine.

Remark 1. We can express p^* in terms of $\mathcal{A}$: $p^* = \sup\{y : (y, \mathbf{0}, \mathbf{0}) \in \mathcal{A}\}$.

Now define

$$\mathcal{B} := \left\{ (y, \mathbf{0}, \mathbf{0}) \in \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^J : y > p^* \right\}.$$

We claim that $\mathcal{A} \cap \mathcal{B} \neq \varnothing$. To see this, suppose not so that there is some $(y, \mathbf{h}, \mathbf{g}) \in \mathcal{A} \cap \mathcal{B}$. That $(y, \mathbf{h}, \mathbf{g}) \in \mathcal{B}$ implies $\mathbf{h} = \mathbf{0}$ and $\mathbf{g} = \mathbf{0}$ and $y > p^*$. That $(y, \mathbf{h}, \mathbf{g}) \in \mathcal{A}$ implies there is an $\mathbf{x} \in \mathbb{R}^d$ such that $A\mathbf{x} = \mathbf{0}$, $g_j(\mathbf{x}) \geq 0$ for all $j \in \{1, 2, \ldots, J\}$, and $f(\mathbf{x}) \geq y$. That is, there exists a feasible $\mathbf{x}$ such that $f(\mathbf{x}) > p^*$ which contradicts that p^* as the value of the primal problem.

We can appeal to the separating hyperplane theorem to obtain a tuple $(c, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) \neq \mathbf{0}$ and α such that

$$(y, \mathbf{h}, \mathbf{g}) \in \mathcal{A} \Rightarrow (y, \mathbf{h}, \mathbf{g}) \cdot (c, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) = cy + \tilde{\boldsymbol{\mu}}^\top \mathbf{h} + \tilde{\boldsymbol{\lambda}}^\top \mathbf{g} \leq \alpha \tag{9}$$

and

$$(y, \mathbf{h}, \mathbf{g}) \in \mathcal{B} \Rightarrow (y, \mathbf{h}, \mathbf{g}) \cdot (c, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) = cy + \tilde{\boldsymbol{\mu}}^\top \mathbf{h} + \tilde{\boldsymbol{\lambda}}^\top \mathbf{g} \geq \alpha. \tag{10}$$

We first argue that $\tilde{\boldsymbol{\lambda}} \geq \mathbf{0}$ and $c \geq 0$. To see this, if $c < 0$, then observe that cy can be made arbitrarily large over $\mathcal{A}$ to violate (9). Similarly, if $\lambda_j < 0$ for some $j \in \{1, 2, \ldots, J\}$, we can also make $\lambda_j g_j$ arbitrarily large over $\mathcal{A}$ to violate (9). Now, (10) means that $cy \geq \alpha$ for all $y > p^*$ so that $cp^* \geq \alpha$. Together with (9), we conclude that

$$\sum_{j=1}^{J} \tilde{\lambda}_j g_j(\mathbf{x}) + \tilde{\boldsymbol{\mu}}^\top (A\mathbf{x} - \mathbf{b}) + cf(\mathbf{x}) \leq \alpha \leq cp^* \;\forall \mathbf{x} \in \mathbb{R}^d$$

Suppose $c > 0$. Then, we can divide above by c to obtain

$$L\left(\mathbf{x}, \frac{\tilde{\boldsymbol{\mu}}}{c}, \frac{\tilde{\boldsymbol{\lambda}}}{c}\right) \leq p^* \;\forall \mathbf{x} \in \mathbb{R}^d.$$

By maximising over $\mathbf{x}$, it follows then that $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq p^*$ where $\boldsymbol{\mu} := \frac{\tilde{\boldsymbol{\mu}}}{c}$ and $\boldsymbol{\lambda} := \frac{\tilde{\boldsymbol{\lambda}}}{c}$. By weak duality, we have $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq p^*$ and so we must in fact have $q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = p^*$. That is, strong duality holds and there the dual optimum is attained.

Now suppose $c = 0$. Then,

$$\sum_{j=1}^{J} \tilde{\lambda}_j g_j(\mathbf{x}) + \tilde{\boldsymbol{\mu}}^\top (A\mathbf{x} - \mathbf{b}) \leq 0 \;\forall \mathbf{x} \in \mathbb{R}^d. \tag{11}$$

Choose $\tilde{\mathbf{x}}$ that satisfies the Slater's condition in which case we must have $\sum_{j=1}^{J} \tilde{\lambda}_j g_j(\tilde{\mathbf{x}}) \leq 0$. Since $g_j(\tilde{\mathbf{x}}) > 0$ for all $j \in \{1, 2, \ldots, J\}$ and $\tilde{\lambda}_j \geq 0$, we must have $\tilde{\lambda}_j = 0$. However, recall that $(c, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\lambda}}) \neq \mathbf{0}$ so that we must have $\tilde{\boldsymbol{\mu}} \neq \mathbf{0}$. The inequality (11) implies $\tilde{\boldsymbol{\mu}}^\top (A\mathbf{x} - \mathbf{b}) \leq 0$ for all $\mathbf{x} \in \mathbb{R}^d$. But that $\tilde{\mathbf{x}}$ satisfies $A\tilde{\mathbf{x}} = \mathbf{b}$ so that $\tilde{\boldsymbol{\mu}}^\top (A\tilde{\mathbf{x}} - \mathbf{b}) = 0$ and there exists $\mathbf{x} \in \mathbb{R}^d$ (increase/decrease one coordinate) such that $\tilde{\boldsymbol{\mu}}^\top (A\mathbf{x} - \mathbf{b}) > 0$ unless $\tilde{\boldsymbol{\mu}}^\top A = \mathbf{0}$. But $\tilde{\boldsymbol{\mu}}^\top A = \mathbf{0}$ would contradict that $\mathrm{rank}(A) = K$.