MATH3027: Optimization (UK 22/23)

Week 11: Duality

Prof. Richard Wilkinson
School of Mathematical Sciences
University of Nottingham, United Kingdom
Please send any comments or mistakes to
r.d.wilkinson@nottingham.ac.uk

This week we conclude our study of constrained optimization problems by discussing the concept of duality. We will show that for certain convex optimization problems, it is possible to formulate an alternative problem, known as the dual, which in some cases cases can be simpler to solve than the original. After studying the structure of the dual problem and its links with the original (primal) problem, we will discuss three relevant instances where duality is useful.

Motivating example

Consider

$$\begin{aligned} f^* = \min \quad & x_1^2 + x_2^2 + 2x_1 \\ \text{s.t.} \quad & x_1 + x_2 = 0. \end{aligned} \tag{P}$$

We can find a lower bound on the value of the problem by considering the unconstrained problem[1]

$$(P_0) \quad f_0^* = \min \quad x_1^2 + x_2^2 + 2x_1$$

Then we can see that we must have $f_0^* \leq f^*$. In fact for this problem, it easy to show that $f^* = -\frac{1}{2}$ and $f_0^* = -1$. In order to find other lower bounds, we can use the following trick: firstly, note that

[1] Note that it is typically much easier to solve unconstrained optimization problems

$$\begin{aligned} f^* = \min \quad & x_1^2 + x_2^2 + 2x_1 + \mu(x_1 + x_2) \\ \text{s.t.} \quad & x_1 + x_2 = 0. \end{aligned}$$

Now we can eliminate the equality constraint and consider the family of unconstrained problems

$$(P_\mu) \quad f_\mu^* = \min \quad x_1^2 + x_2^2 + 2x_1 + \mu(x_1 + x_2)$$

We have that

$$f_\mu^* \leq f^* \text{ for all } \mu.$$

With a bit of work you can show that $f_\mu^* = -\frac{\mu^2}{2} - \mu - 1$. What interests us is the best (i.e., largest) lower bound obtained by this technique. This best lower bound is found by solving the *dual* problem.

$$\max\{q(\mu) : \mu \in \mathbb{R}\}. \tag{Dual}$$

By construction, the dual problem provides a lower bound on the original problem (P), which we will call the *primal* problem. In this case, the optimal value of the dual problem is attained at $\mu = -1$ and the corresponding optimal value of the dual problem is $f_{-1}^* = -\frac{1}{2} = f^*$, i.e., in this case the best lower bound is equal to the optimal value f^* of the primal problem. When this occurs, we refer to this property as 'strong duality'.

We now discuss the construction of the dual problem more formally.

The Primal and Dual Problems

Consider the problem

$$\begin{aligned} f^* := \min\ & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \\ & h_j(\mathbf{x}) = 0, j = 1, 2, \ldots, p, \\ & \mathbf{x} \in X, \end{aligned} \tag{Primal}$$

where $f, g_i, h_j (i = 1, 2, \ldots, m, j = 1, 2, \ldots, p)$ are functions defined on the set $X \subseteq \mathbb{R}^n$. This is the "usual" optimization problem, and we will refer to it as the **primal** problem. As discussed last week, the Lagrangian associated with this problem is

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \mu_j h_j(\mathbf{x}) \quad \left(\mathbf{x} \in X, \boldsymbol{\lambda} \in \mathbb{R}_+^m, \boldsymbol{\mu} \in \mathbb{R}^p\right).$$

The **dual** objective function $q : \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ is defined to be

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{1}$$

The domain of the dual objective function is

$$\operatorname{dom}(q) = \left\{ (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^m_+ \times \mathbb{R}^p : q(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty \right\} .$$

The **dual problem** is given by

$$\begin{aligned} q^* &:= \max \, q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \operatorname{dom}(q) \end{aligned} \tag{Dual}$$

The solution to the dual problem provides a lower bound on the solution of the primal problem, and in many cases the two solutions are equal. Thus in many optimization problems it is useful to solve the dual problem instead of the primal problem. When and why this is a good idea is an open-ended question that we shall explore this week, illustrating the method with some relevant examples.

We begin by stating some properties of the dual problem.

Theorem. *Consider the primal problem* (Primal) *with* $f, g_i, h_j (i = 1, 2, \ldots, m, \, j = 1, 2, \ldots, p)$ *functions defined on* $X \subseteq \mathbb{R}^n$*, and let* q *be the dual function defined in Equation (1). Then:*

a) $\operatorname{dom}(q)$ *is a convex set,*

b) q *is a concave function over* $\operatorname{dom}(q)$*.*

Proof. (a) Take $\left(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1\right), \left(\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) \in \operatorname{dom}(q)$ and $\alpha \in [0, 1]$. Then

$$\begin{aligned} \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\mu}_1\right) &> -\infty , \\ \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) &> -\infty . \end{aligned}$$

Therefore, since the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is affine w.r.t. $\boldsymbol{\lambda}, \boldsymbol{\mu}$

$$\begin{aligned} q\left(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\boldsymbol{\mu}_1 + (1-\alpha)\boldsymbol{\mu}_2\right) &= \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\boldsymbol{\mu}_1 + (1-\alpha)\boldsymbol{\mu}_2\right) \\ &= \min_{\mathbf{x} \in X} \left\{ \alpha L\left(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\mu}_1\right) + (1-\alpha) L\left(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) \right\} \\ &\geq \alpha \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}_1\right) + (1-\alpha) \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) \\ &= \alpha q\left(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1\right) + (1-\alpha) q\left(\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) \\ &> -\infty \end{aligned}$$

Hence, $\alpha\left(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1\right) + (1-\alpha)\left(\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2\right) \in \operatorname{dom}(q)$, and the convexity of $\operatorname{dom}(q)$ is established.
(b) $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is an affine function w.r.t. $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. In particular, it is a concave function w.r.t. $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. Hence, since q is the minimum of concave functions, it must be concave. Maximizing a concave function is equivalent to minimizing a convex function, and so the dual problem is also a convex optimization problem. □

Weak and Strong Duality

A first important consequence for optimization is the weak duality theorem, which establishes a lower bound for the optimal primal value using the optimal dual value.

Theorem (Weak Duality Theorem)**.** *Consider the primal problem* (Primal) *and its dual problem* (Dual)*. Then*

$$q^* \leq f^*$$

where f^*, q^* are the optimal primal and dual values respectively.

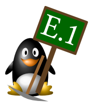
Proof. The feasible set of the primal problem is

$$S = \left\{ \mathbf{x} \in X : g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0, i = 1, 2, \ldots, m, j = 1, 2, \ldots, p \right\} .$$

For any $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \operatorname{dom}(q)$ we have

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &\leq \min_{\mathbf{x} \in S} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \min_{\mathbf{x} \in S} \left\{ f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \right\} \\ &\leq \min_{\mathbf{x} \in S} f(\mathbf{x}) = f^* \end{aligned}$$

where the last inequality follows as $\lambda_i \geq 0$ and $g_i(\mathbf{x}) \leq 0$, $h_j(\mathbf{x}) = 0$ for $\mathbf{x} \in S$. Taking the maximum over $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \operatorname{dom}(q)$, gives the required result. □

Example: 

$$\begin{array}{ll} \min & x_1^2 - 3x_2^2 \\ \text{s.t.} & x_1 = x_2^3 \end{array}$$

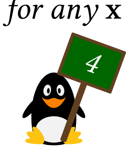
While the weak duality theorem is useful for obtaining a lower bound for the optimal value of the primal problem, a more powerful result can be proven known as **strong duality**. For this, we need a nonlinear variant of Farkas' Lemma, which we now state.

Theorem (Supporting Hyperplane Theorem)**.** *Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{y} \notin C$. Then there exists $\mathbf{0} \neq \mathbf{p} \in \mathbb{R}^n$ such that*

$$\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^T\mathbf{y} \text{ for any } \mathbf{x} \in C .$$

Theorem (Separation of Two Convex Sets)**.** *Let $C_1, C_2 \subseteq \mathbb{R}^n$ be two nonempty convex sets such that $C_1 \cap C_2 = \emptyset$. Then there exists $\mathbf{0} \neq \mathbf{p} \in \mathbb{R}^n$ for which*

$$\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^T\mathbf{y} \text{ for any } \mathbf{x} \in C_1, \mathbf{y} \in C_2 .$$



Theorem (Nonlinear Farkas Lemma). *Let $X \subseteq \mathbb{R}^n$ be a convex set and let $f, g_1, g_2, \ldots, g_m$ be convex functions over X. Assume that there exists $\hat{x} \in X$ such that*

$$g_1(\hat{\mathbf{x}}) < 0, g_2(\hat{\mathbf{x}}) < 0, \ldots, g_m(\hat{\mathbf{x}}) < 0\,. \quad \textit{(Slater's condition)}$$

Let $c \in \mathbb{R}$. Then the following two claims are equivalent:

a) The following implication holds:

$$\mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \Rightarrow f(\mathbf{x}) \geq c\,.$$

b) There exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that

$$\min_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right\} \geq c\,.$$

With these technical results, we are in position to state the strong duality result.

Theorem (Strong Duality of Convex Problems with Inequality Constraints). *Consider the optimization problem*

$$\begin{aligned} f^* = \min\ & f(\mathbf{x}) \\ \textit{s.t.}\ \ & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\ & \mathbf{x} \in X \end{aligned}$$

where X is a convex set and $f, g_i, i = 1, 2, \ldots, m$ are convex functions over X. Suppose that Slater's condition holds, i.e., there exists $\hat{\mathbf{x}} \in X$ for which $g_i(\hat{\mathbf{x}}) < 0, i = 1, 2, \ldots, m$. If this problem has a finite optimal value, then

a) the optimal value of the dual problem is attained.

b) the primal and dual problems have the same optimal value, $f^* = q^*$.

Proof. Since $f^* > -\infty$ is the optimal value of the primal problem, it follows that the following implication holds:

$$\mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \Rightarrow f(\mathbf{x}) \geq f^*\,.$$

By the nonlinear Farkas Lemma, there exists $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_m \geq 0$ such that

$$q(\tilde{\lambda}) = \min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sum_{j=1}^{m} \tilde{\lambda}_j g_j(\mathbf{x}) \right\} \geq f^*\,.$$

By the weak duality theorem,

$$q^* \geq q(\tilde{\boldsymbol{\lambda}}) \geq f^* \geq q^*\,.$$

Hence, $f^* = q^*$ and $\tilde{\boldsymbol{\lambda}}$ is an optimal solution of the dual problem. $\square$

The result above indicates that under the convexity assumptions of the theorem, it is possible to obtain the solution of the primal problem by solving its dual.

Another useful result regarding dual problems is the following. It shows that we can derive the complementary slackness conditions under the assumption that $q^* = f^*$ (i.e. without assuming convexity).

Theorem (Complementary Slackness Conditions)**.** *Consider the optimization problem*

$$f^* := \min \left\{ f(\mathbf{x}) : g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m, \ \mathbf{x} \in X \right\} ,$$

and assume that $f^* = q^*$ where q^* is the optimal value of the dual problem. Let $\mathbf{x}^*, \boldsymbol{\lambda}^*$ be feasible solutions of the primal and dual problems. Then $\mathbf{x}^*, \boldsymbol{\lambda}^*$ are optimal solutions of the primal and dual problems iff

$$\begin{aligned} \mathbf{x}^* &\in \underset{\mathbf{x} \in X}{\operatorname{argmin}}\, L\left(\mathbf{x}, \boldsymbol{\lambda}^*\right) \\ \lambda_i^* g_i\left(\mathbf{x}^*\right) &= 0, i = 1, 2, \ldots, m \end{aligned}$$

Proof. We have

$$q^* = q\left(\lambda^*\right) = \min_{\mathbf{x} \in X} L\left(\mathbf{x}, \lambda^*\right) \leq L\left(\mathbf{x}^*, \lambda^*\right) = f\left(\mathbf{x}^*\right) + \sum_{i=1}^{m} \lambda_i^* g_i\left(\mathbf{x}^*\right) \leq f\left(\mathbf{x}^*\right) = f^* .$$

Since $f^* = q^*$, the inequalities are equalities, and so we have $\mathbf{x}^* \in \arg\min L(\mathbf{x}, \lambda^*)$ and $\sum \lambda_i g_i(\mathbf{x}^*) = 0$. As $\lambda \geq 0$ and $g_i(x) \leq 0$ we must have $\lambda_i g_i(\mathbf{x}) = 0$. □

E.2

Find a dual problem to the convex problem

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + x_1 x_2 - x_1 \\ \text{s.t.} & 3x_1 + x_2 \leq 1. \end{array}$$

Find the optimal solution of both the dual and primal problems.

Note that Slater's condition is important, without it, strong duality may not apply, and we may have a duality gap:

E.3

What goes wrong when we find the dual of this problem?

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \textbf{s.t.} & x_2^2 \leq 0 \, . \end{array}$$

Duffin's Duality Gap. Consider the problem

$$\min\left\{e^{-x_2} : \sqrt{x_1^2 + x_2^2} - x_1 \leq 0\right\} .$$

The feasible set is in fact $F = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\} \implies f^* = 1$. Note that Slater's condition does not hold here.

The Lagrangian is given by

$$L\left(x_1, x_2, \lambda\right) = e^{-x_2} + \lambda\left(\sqrt{x_1^2 + x_2^2} - x_1\right), \quad (\lambda \geq 0) .$$

The dual objective is given by

$$q(\lambda) = \min_{x_1, x_2} L\left(x_1, x_2, \lambda\right) .$$

First of all, note that $L(\mathbf{x}, \lambda) \geq 0$ for all $\mathbf{x}$ and hence $q(\lambda) \geq 0$. We will show that the minimum is 0 regardless of the value of λ, i.e., that $q(\lambda) = 0$ for all λ.

For any $\varepsilon > 0$, if we choose $x_2 = -\log \varepsilon$ and $x_1 = \frac{x_2^2 - \varepsilon^2}{2\varepsilon}$, then

$$\begin{aligned}
\sqrt{x_1^2 + x_2^2} - x_1 &= \sqrt{\frac{\left(x_2^2 - \varepsilon^2\right)^2}{4\varepsilon^2} + x_2^2} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\
&= \sqrt{\frac{\left(x_2^2 + \varepsilon^2\right)^2}{4\varepsilon^2}} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\
&= \frac{x_2^2 + \varepsilon^2}{2\varepsilon} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \varepsilon .
\end{aligned}$$

Hence, $L\left(x_1, x_2, \lambda\right) = e^{-x_2} + \lambda\left(\sqrt{x_1^2 + x_2^2} - x_1\right) = \varepsilon + \lambda\varepsilon = (1 + \lambda)\varepsilon$. This implies that $q(\lambda) = 0$ for all $\lambda \geq 0$. Therefore the dual problem is $\max\{0 : \lambda \geq 0\}$ and so $q^* = 0$. Thus in this case the duality gap is

$$f^* - q^* = 1 \implies \text{ a duality gap of } 1 .$$

This is a result of the fact that Slater's condition is not satisfied here.

We conclude these results with a more general duality theorem including convex affine inequality and equality contraints.

Theorem (General Strong Duality Theorem). *Consider the optimization problem*

$$\begin{aligned}
f^* = \min \;& f(\mathbf{x}) \\
\text{s.t. } g_i(\mathbf{x}) &\leq 0, \quad i = 1, 2, \ldots, m \\
h_j(\mathbf{x}) &\leq 0, \quad j = 1, 2, \ldots, p \\
s_k(\mathbf{x}) &= 0, \quad k = 1, 2, \ldots, q \\
\mathbf{x} &\in X ,
\end{aligned}$$

where X is a convex set and $f, g_i, i = 1, 2, \ldots, m$ are convex functions over X. The functions h_j, s_k are affine functions. Suppose that there exists[2] $\hat{\mathbf{x}} \in \operatorname{int}(X)$ *for which* $g_i(\hat{\mathbf{x}}) < 0$, $h_j(\hat{\mathbf{x}}) \leq 0$, *and* $s_k(\hat{\mathbf{x}}) = 0$. *Then if the problem has a finite optimal value, then the optimal value of the dual problem*

$$q^* = \max\{q(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) : (\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) \in \operatorname{dom}(q)\}$$

where

$$q(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in X} \left[f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \eta_j h_j(\mathbf{x}) + \sum_{k=1}^{q} \mu_k s_k(\mathbf{x}) \right]$$

is attained, and $f^* = q^*$.

Example. E.4 (Check the details) Consider the problem

$$\begin{aligned} \min x_1^3 + x_2^3 \\ x_1 + x_2 \geq 1 \\ x_1, x_2 \geq 0\,. \end{aligned}$$

We will show that a problem can have different dual formulations with different duality gaps, depending on how we choose X. Through the usual KKT conditions, we can easily show that $\left(\frac{1}{2}, \frac{1}{2}\right)$ is the optimal solution of the primal problem with an optimal value $f^* = \frac{1}{4}$.

One way to construct a dual problem is to take $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$. The primal problem is then

$$\min \left\{ x_1^3 + x_2^3 : x_1 + x_2 \geq 1, (x_1, x_2) \in X \right\}.$$

Strong duality holds for the problem[3] and hence we have $q^* = \frac{1}{4}$.

A second dual is constructed by taking $X = \mathbb{R}^2$. In this case, the objective function $f(x_1, x_2) = x_1^3 + x_2^3$ is not convex, implying that strong duality is not necessarily satisfied. The Lagrangian in this case is given by

$$L\left(x_1, x_2, \lambda, \eta_1, \eta_2\right) = x_1^3 + x_2^3 - \lambda\left(x_1 + x_2 - 1\right) - \eta_1 x_1 - \eta_2 x_2\,.$$

We can see that $q\left(\lambda, \eta_1, \eta_2\right) = -\infty$ for all $\left(\lambda, \eta_1, \eta_2\right) \Rightarrow q^* = -\infty$, and the duality gap is infinite.

This illustrates that it is important to make a good choice of X to obtain useful information about the solution of the primal problem.

---

[2] This is the generalized Slater's condition we saw last week.

[3] Why is the problem convex?

Three Important Examples of Duality Use

Linear Programming

Consider the linear programming problem

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \end{array}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. For linear programmes, if the primal has a finite solution then strong duality always applies. The Lagrangian is given by

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \left(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}\right)^T \mathbf{x} - \mathbf{b}^T\boldsymbol{\lambda},$$

and the dual objective function is

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\mathbf{x}\in\mathbb{R}^n} \left(\mathbf{c} + \mathbf{A}^T\lambda\right)^T \mathbf{x} - \mathbf{b}^T\boldsymbol{\lambda} = \begin{cases} -\mathbf{b}^T\boldsymbol{\lambda} & \text{if } \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} = 0 \\ -\infty & \text{else.} \end{cases}$$

Therefore, the dual problem is formulated as

$$\begin{array}{ll} \max & -\mathbf{b}^T\boldsymbol{\lambda} \quad (\equiv \min \mathbf{b}^T\boldsymbol{\lambda}) \\ \text{s.t.} & \mathbf{A}^T\boldsymbol{\lambda} = -\mathbf{c} \\ & \boldsymbol{\lambda} \geq \mathbf{0}\,. \end{array}$$

E.5

Find the dual problem of

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Strictly Convex Quadratic Programming

Consider the strictly convex quadratic programming problem

$$\begin{array}{ll} \min & x^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}\,, \end{array}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive definite, $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, and $\mathbf{b} \in \mathbb{R}^m$. The Lagrangian (recall $\boldsymbol{\lambda} \in \mathbb{R}^m_+$) is given by[4]:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} + 2\boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\left(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{f}\right)^T \mathbf{x} - 2\mathbf{b}^T\boldsymbol{\lambda}\,.$$

[4] The factor of 2 doesn't matter here - it just simplifies things later.

The minimizer of the Lagrangian is attained at a stationary point which is the solution to $\nabla_x L = 0$, which implies $\mathbf{x}^* = -\mathrm{Q}^{-1}\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right)$. With this, we find the dual objective,

$$
\begin{aligned}
q(\boldsymbol{\lambda}) &= L\left(\mathbf{x}^*, \boldsymbol{\lambda}\right) \\
&= \left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right)^T \mathrm{Q}^{-1}\mathrm{Q}\mathrm{Q}^{-1}\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right) - 2\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right)^T \mathrm{Q}^{-1}\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right) - 2\mathbf{b}^T\boldsymbol{\lambda} \\
&= -\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right)^T \mathrm{Q}^{-1}\left(\mathbf{f} + \mathrm{A}^T\boldsymbol{\lambda}\right) - 2\mathbf{b}^T\boldsymbol{\lambda} \\
&= -\boldsymbol{\lambda}^T\mathrm{A}\mathrm{Q}^{-1}\mathrm{A}^T\boldsymbol{\lambda} - 2\mathbf{f}^T\mathrm{Q}^{-1}\mathrm{A}^T\boldsymbol{\lambda} - \mathbf{f}^T\mathrm{Q}^{-1}\mathbf{f} - 2\mathbf{b}^T\boldsymbol{\lambda} \\
&= -\boldsymbol{\lambda}^T\mathrm{A}\mathrm{Q}^{-1}\mathrm{A}^T\boldsymbol{\lambda} - 2\left(\mathrm{A}\mathrm{Q}^{-1}\mathbf{f} + \mathbf{b}\right)^T\boldsymbol{\lambda} - \mathbf{f}^T\mathrm{Q}^{-1}\mathbf{f}\,.
\end{aligned}
$$

The resulting dual problem is then $\max\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \geq 0\}$. This is another convex optimization problem, but the dual has a simpler feasible set than the primal problem and can be solved using the projected gradient methods described previously.

In the computer lab you will find $P_S(\mathbf{y})$ where $S = \{\mathbf{x} \in R^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ using duality methods.

Computing the Orthogonal Projection onto the Unit Simplex

Given a vector $\mathbf{y} \in \mathbb{R}^n$, we would like to compute the orthogonal projection of the vector $\mathbf{y}$ onto Δ_n, i.e., $P_{\Delta_n}(\mathbf{y})$. The corresponding optimization problem is

$$
\begin{array}{ll}
\min_{\mathbf{x}} & \|\mathbf{x} - \mathbf{y}\|^2 \\
\text{s.t.} & \mathbf{e}^T\mathbf{x} = 1 \\
& \quad \mathbf{x} \geq 0\,.
\end{array}
$$

where $\mathbf{e} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}$. We will associate a Lagrange multiplier $\lambda \in \mathbb{R}$ to the linear equality constraint $\mathbf{e}^T\mathbf{x} = 1$ and obtain the Lagrangian function

$$
\begin{aligned}
L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{y}\|^2 + 2\lambda\left(\mathbf{e}^T\mathbf{x} - 1\right) &= \|\mathbf{x}\|^2 - 2(\mathbf{y} - \lambda\mathbf{e})^T\mathbf{x} + \|\mathbf{y}\|^2 - 2\lambda \\
&= \sum_{j=1}^{n}\left(x_j^2 - 2\left(y_j - \lambda\right)x_j\right) + \|\mathbf{y}\|^2 - 2\lambda\,.
\end{aligned}
$$

This problem is separable with respect to the variables x_j and hence the optimal x_j is the solution to the one-dimensional problem

$$
\min_{x_j \geq 0}\left[x_j^2 - 2\left(y_j - \lambda\right)x_j\right]
$$

The optimal solution to the above problem is given by

$$
\begin{aligned}
x_j &= \begin{cases} y_j - \lambda, & y_j \geq \lambda \\ 0 & \text{otherwise} \end{cases} \\
&= \left[y_j - \lambda\right]_+
\end{aligned}
$$

and the optimal value is $-\left[y_j-\lambda\right]_+^2$. The dual problem is therefore

$$\max_{\lambda\in\mathbb{R}}\left\{g(\lambda)\equiv-\sum_{j=1}^{n}\left[y_j-\lambda\right]_+^2-2\lambda+\|\mathbf{y}\|^2\right\}.$$

By the basic properties of dual problems, the dual objective function is concave. In order to actually solve the dual problem, we note that

$$\lim_{\lambda\to\infty}g(\lambda)=\lim_{\lambda\to-\infty}g(\lambda)=-\infty$$

i.e., g is a coercive and differentiable function. It follows that there exists an optimal solution to the dual problem attained at a point λ in which

$$g'(\lambda)=0\,,$$

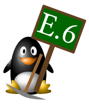
meaning that

$$\sum_{j=1}^{n}\left[y_j-\lambda\right]_+=1\,.$$

The function $h(\lambda)=\sum_{j=1}^{n}\left[y_j-\lambda\right]_+-1$ is a nonincreasing function over $\mathbb{R}$ and is in fact strictly decreasing over $\left(-\infty,\max_j y_j\right]$. We need to find a root of h i.e., a solution to $h(\lambda)=0$. By denoting $y_{\max}=\max_{j=1,2,\ldots,n}y_j$, and $y_{\min}=\min_{j=1,2,\ldots,n}y_j$, we have

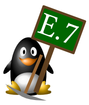
$$\begin{aligned}h\left(y_{\max}\right)&=-1\\ h\left(y_{\min}-\frac{2}{n}\right)&=\sum_{j=1}^{n}y_j-ny_{\min}+2-1>0\,,\end{aligned}$$

and we can therefore invoke a bisection procedure to find the unique root λ of the function h over the interval $\left[y_{\min}-\frac{2}{n},y_{\max}\right]$ and then define $P_{\Delta_n}(\mathbf{y})=[\mathbf{y}-\lambda\mathbf{e}]_+$.



Implement this in R and find the orthogonal projection of $(-1,1,0.3)$ onto Δ_n.

Bonus questions



Find a dual problem to the problem

$$\begin{aligned}f^*:=\min\ & x_1-4x_2+x_3^4\\ \text{s.t.}\quad & x_1+x_2+x_3^2\le 2\\ & x_1\ge 0\\ & x_2\ge 0.\end{aligned}\tag{Primal}$$

Solve the dual and primal problems.

E.8 Consider the optimization problem

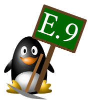
$$\begin{aligned} \min \quad & \sum_{j=1}^{n} \frac{c_j}{x_j} \\ \text{s.t.} \quad & \mathbf{a}^\top \mathbf{x} \le b \\ & \mathbf{x} \ge 0. \end{aligned} \tag{Primal}$$

where $\mathbf{a}, \mathbf{c} \in \mathbb{R}^n_{++}$ and $b \in \mathbb{R}_{++}$.

- Find a dual problem with a single dual decision variable.
- Solve the dual and primal problems.

Checklist

The idea of this checklist is to help you to self-evaluate your progress and understanding of the subject, and to give you some guidance on where to focus. If you can tick all the boxes it means you're doing alright, otherwise you need to study a bit more, grab a book, watch the videos, or seek help from classmates, the lecturers, or the demonstrators. Try to fill as many gaps as quickly as possible.

And remember to do the 

's!

Learning Outcome	Check
I understand what is the dual problem.	
I understand the definition of weak and strong duality.	
I understand the concept of duality gap.	
I can work the examples in the lecture notes and derive the dual of optimization problems and appreciate why sometimes it's simpler to solve.	