Optimization (D2) **DPK**

Easter 2005

1. LAGRANGIAN METHODS

1.1 Constrained optimization

The general problem of optimization under constraints that we will consider in this course may be expressed as

$$\mathrm{P}: \quad \text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) = \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

where

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is known as the **objective function** of the problem and the components of its argument $\boldsymbol{x} = (x_1, \ldots x_n)^\top$ are the variables to be chosen to optimize the objective function subject to the constraints;

(ii) $X \subseteq \mathbb{R}^n$ is the **constraint region**;

(iii) $g : \mathbb{R}^n \to \mathbb{R}^m$ defines m functional constraints; and

(iv) $\boldsymbol{b} = (b_1, \ldots, b_m)^\top$ is a fixed element of $\mathbb{R}^m$, sometimes known, prosaically, as the **right-hand side** of the constraints.

The convention throughout will be that all vectors are column vectors. While we will formulate the general problem as a minimization one may move between maximization and minimization problems by observing that maximizing f is equivalent to minimizing $-f$. The region X, which is a subset of $\mathbb{R}^n$, is usually a set such as the non-negative orthant $X = \{\boldsymbol{x} : \ \boldsymbol{x} \geqslant 0\}$, and in some problems the functional constraints may take the form of inequalities $g(\boldsymbol{x}) \leqslant \boldsymbol{b}$; these inequality constraints may be turned into equality constraints by the addition of non-negative **slack** variables $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$ to give

$$g(\boldsymbol{x}) + \boldsymbol{z} = \boldsymbol{b}, \quad \boldsymbol{z} \geqslant 0.$$

It can be seen from this that there may be a degree of arbitrariness in how the constraints on $\boldsymbol{x}$ are divided between functional constraints and the constraint region but it is usually clear from the context of the problem how each should be formulated.

The set $X_{\boldsymbol{b}} = \{\boldsymbol{x} : \boldsymbol{x} \in X, \ g(\boldsymbol{x}) = \boldsymbol{b}\}$ is known as the set of **feasible** solutions of the problem P, so the constrained problem is to find a feasible $\boldsymbol{x}^* \in X_{\boldsymbol{b}}$ that minimizes the function f over $X_{\boldsymbol{b}}$, that is, $f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X_{\boldsymbol{b}}$; such a $\boldsymbol{x}^*$ is said to be **optimal** for the problem P. Techniques of unconstrained optimization are usually straightforward, to minimize f the first step would be to set the gradient of f equal to zero; when the problem is constrained such an approach may not be immediately helpful as the optimum will not necessarily occur when the gradient of f is zero. For example, consider the case when $f(x) = x^2$ and $X = \{x : x \geqslant 2\}$; the minimum trivially occurs when $x = 2$ but $f'(2) \neq 0$. The Lagrangian approach is a powerful method which enables many constrained problems to be turned into unconstrained optimization problems which can then be tackled by conventional means.

1.2 The Lagrangian Sufficiency Theorem

The **Lagrangian** for the problem P is defined to be the function

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top (\boldsymbol{b} - g(\boldsymbol{x}))$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$; here, $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the components of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top$ are known as the **Lagrange multipliers** of the problem. The following very simple result is central to the Lagrangian technique.

Theorem 1.1 **(Lagrangian Sufficiency Theorem)** *Suppose that there exist $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}^*$ such that*

$$L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{\lambda}^*) \quad \textit{for all} \quad \boldsymbol{x} \in X,$$

and $\boldsymbol{x}^*$ is feasible for P, *then $\boldsymbol{x}^*$ is optimal for* P.

Proof. For any $\boldsymbol{x} \in X_{\boldsymbol{b}}$,

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top (\boldsymbol{b} - g(\boldsymbol{x})) = f(\boldsymbol{x}),$$

which shows that

$$f(\boldsymbol{x}^*) = L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{\lambda}^*), \quad \text{for all } \boldsymbol{x} \in X;$$

but $L(\boldsymbol{x}, \boldsymbol{\lambda}^*) = f(\boldsymbol{x})$ for $\boldsymbol{x} \in X_{\boldsymbol{b}}$ and hence $f(\boldsymbol{x}^*) \leqslant f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X_{\boldsymbol{b}}$, which is the result. $\square$

The Lagrangian Sufficiency Theorem leads to the following procedure which may be used to solve a wide class of constrained optimization problems.

Lagrangian Method for Constrained Optimization

1. For each $\boldsymbol{\lambda}$ for which $\inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}) > -\infty$, choose $\boldsymbol{x}^*(\boldsymbol{\lambda}) \in X$ so that

   $$L(\boldsymbol{x}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}).$$

   That is, for each $\boldsymbol{\lambda}$, we minimize $L(\boldsymbol{x}, \boldsymbol{\lambda})$ over $\boldsymbol{x}$ in X, unconstrained by the functional constraints $g(\boldsymbol{x}) = \boldsymbol{b}$, to get the minimizing $\boldsymbol{x}^*(\boldsymbol{\lambda})$. We may exclude any $\boldsymbol{\lambda}$ for which $\inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}) = -\infty$ for such $\boldsymbol{\lambda}$ cannot satisfy the conditions of Theorem 1.1.

2. Now choose $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ so that $\boldsymbol{x}^* = \boldsymbol{x}^*(\boldsymbol{\lambda}^*) \in X_{\boldsymbol{b}}$ (in other words, to satisfy $g(\boldsymbol{x}^*(\boldsymbol{\lambda}^*)) = \boldsymbol{b}$ so that $\boldsymbol{x}^*$ is feasible for the problem P); then $\boldsymbol{x}^*$, $\boldsymbol{\lambda}^*$ satisfy the conditions of Theorem 1.1 and hence $\boldsymbol{x}^*$ is optimal for P.

Example 1.2 *Use of the Lagrangian Method*

Consider the problem

$$\begin{array}{ll} \text{minimize} & 2\ln(1/x_1) + 3\ln(1/x_2) + \ln(1/x_3) \\ \text{subject to} & 3x_1 + 2x_2 + x_3 = 1, \quad x_1 > 0,\ x_2 > 0,\ x_3 > 0. \end{array}$$

The Lagrangian is

$$L(\boldsymbol{x}, \lambda) = 2\ln(1/x_1) + 3\ln(1/x_2) + \ln(1/x_3) + \lambda(1 - 3x_1 - 2x_2 - x_3),$$

which we have to minimize over the variables $x_1 > 0$, $x_2 > 0$ and $x_3 > 0$. We set the first derivatives of L with respect to each of these to zero to get

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= -\frac{2}{x_1} - 3\lambda = 0 \\ \frac{\partial L}{\partial x_2} &= -\frac{3}{x_2} - 2\lambda = 0 \\ \frac{\partial L}{\partial x_3} &= -\frac{1}{x_3} - \lambda = 0, \end{aligned}$$

which give $x_1^*(\lambda) = -2/(3\lambda)$, $x_2^*(\lambda) = -3/(2\lambda)$, $x_3^*(\lambda) = -1/\lambda$; we need to ensure that this turning point of $L(\boldsymbol{x}, \lambda)$ is a minimum, so consider the Hessian matrix of L (the matrix of second derivatives)

$$\mathcal{H}_L = \left(\frac{\partial^2 L}{\partial x_i \partial x_j}\right)_{ij} = \begin{pmatrix} 2/x_1^2 & 0 & 0 \\ 0 & 3/x_2^2 & 0 \\ 0 & 0 & 1/x_3^2 \end{pmatrix}$$

which shows that these values give a minimum. Substitute $x_i^*(\lambda)$, $i = 1, 2, 3$, into the constraint $3x_1 + 2x_2 + x_3 = 1$

$$3\left(-\frac{2}{3\lambda}\right) + 2\left(-\frac{3}{2\lambda}\right) - \frac{1}{\lambda} = 1,$$

to determine the value of λ which makes these feasible for the problem and we obtain $\lambda = -6$, to give the optimal solution as $x_1^* = 1/9$, $x_2^* = 1/4$ and $x_3^* = 1/6$. $\square$

1.3 Inequality constraints and complementary slackness

When the functional constraints in the problem P are in inequality form so that the problem becomes

$$\mathrm{P}: \quad \text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) \leqslant \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

it may be expressed in the previous form with equality constraints using slack variables as

$$\text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) + \boldsymbol{z} = \boldsymbol{b}, \quad \boldsymbol{x} \in X \text{ and } \boldsymbol{z} \geqslant 0.$$

The Lagrangian now becomes

$$L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top (\boldsymbol{b} - g(\boldsymbol{x}) - \boldsymbol{z})$$

and it must be minimized over $\boldsymbol{x} \in X$ and $\boldsymbol{z} \geqslant 0$. When minimizing over the slack variables $\boldsymbol{z} = (z_1, \ldots, z_m)^\top \geqslant 0$, one must be careful not to set $\partial L/\partial z_i = 0$, and then conclude that $\lambda_i = 0$, since the optimum may occur at the boundary of the region $z_i \geqslant 0$, viz. at $z_i = 0$ when $\partial L/\partial z_i$ is not necessarily zero.

Consider the term in the Lagrangian involving z_i, viz. $-\lambda_i z_i$; if $\lambda_i > 0$ then letting z_i become arbitrarily large shows that this term can be made to approach $-\infty$ which implies that $\inf_{\boldsymbol{x}\in X, \boldsymbol{z}\geqslant 0} L(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\lambda}) = -\infty$. Thus, for a finite minimum of the Lagrangian we require that $\lambda_i \leqslant 0$, in which case the minimum of the term $-\lambda_i z_i$ is 0, since we could take $z_i = 0$. Thus, with inequality constraints in the problem, minimizing the Lagrangian always leads to sign conditions on the Lagrange multipliers, in this case $\boldsymbol{\lambda} \leqslant 0$, and also a joint condition on the Lagrange multipliers and the slack variables in that

$$\lambda_i z_i = 0 \quad \text{for each } i = 1, \ldots, m, \quad \text{or equivalently,} \quad \boldsymbol{\lambda}^\top \boldsymbol{z} = 0.$$

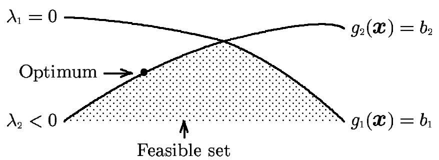
This condition is known as a **complementary slackness** condition; at least one of the variables λ_i and z_i must be zero (at the optimum solution) for each i.

To understand what the complementary slackness condition is saying about the optimal solution to the constrained optimization problem, notice that, at the optimum,

(i) if $z_i > 0$ then necessarily $\lambda_i = 0$; and

(ii) if $\lambda_i < 0$ then necessarily $z_i = 0$.

The statement (i) is saying that if at the optimum the ith constraint is not **tight**, that is, $g_i(\boldsymbol{x}) < b_i$, then the Lagrange multiplier λ_i for that constraint must be zero, so that the term involving the ith constraint is not required in the Lagrangian. Alternatively, (ii) says that if λ_i is not zero, so the term in the Lagrangian for the ith constraint is required, then necessarily $z_i = 0$ so the ith constraint is tight at the optimum.



Sometimes when the optimum to the problem falls on one constraint but inside another (as illustrated in the diagram above and the Example below) the cases $\lambda_i = 0$ and $\lambda_i < 0$ may need to be considered separately for each constraint.

Example 1.3 *Complementary slackness*

Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1 - 3x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 \leqslant 4 \\ & x_1 + x_2 \leqslant 2. \end{aligned}$$

Adding slack variables the problem becomes

$$\begin{aligned} \text{minimize} \quad & x_1 - 3x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 + z_1 = 4 \\ & x_1 + x_2 + z_2 = 2, \quad z_1 \geqslant 0, \ z_2 \geqslant 0. \end{aligned}$$

It is easy to see (by drawing a picture) where the optimum for this problem occurs, but we will use it as an illustration of the methods of this section. The Lagrangian for the problem is

$$L = x_1 - 3x_2 + \lambda_1 \left(4 - x_1^2 - x_2^2 - z_1\right) + \lambda_2 \left(2 - x_1 - x_2 - z_2\right),$$

which must be minimized over $x_1,\ x_2 \in \mathbb{R}$ and $z_1,\ z_2 \geqslant 0$. Minimizing over $z_1 \geqslant 0$ and $z_2 \geqslant 0$, the arguments above show that for a finite minimum we must have $\lambda_1 \leqslant 0$, $\lambda_2 \leqslant 0$ and at the optimum the complementary slackness conditions $\lambda_1 z_1 = 0$ and $\lambda_2 z_2 = 0$ must hold. Minimizing L in x_1 and x_2, we have

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= \ \ 1 - 2\lambda_1 x_1 - \lambda_2 = 0 \\ \frac{\partial L}{\partial x_2} &= -3 - 2\lambda_1 x_2 - \lambda_2 = 0, \end{aligned}$$

which will give a minimum because the Hessian matrix

$$\mathcal{H}_L = \begin{pmatrix} -2\lambda_1 & 0 \\ 0 & -2\lambda_1 \end{pmatrix},$$

is non-negative definite since we have $\lambda_1 \leqslant 0$. Notice that we cannot have $\lambda_1 = 0$, since it would give inconsistent values $\lambda_2 = 1$ and $\lambda_2 = -3$. Suppose that $\lambda_1 < 0$, so necessarily $z_1 = 0$, and

$$x_1 = \frac{1 - \lambda_2}{2\lambda_1}, \quad x_2 = -\frac{3 + \lambda_2}{2\lambda_1};$$

if we further assume that $\lambda_2 < 0$ so that $z_2 = 0$ and substitute into the two constraints to get λ_1 and λ_2 we obtain

$$\left(\frac{1-\lambda_2}{2\lambda_1}\right)^2 + \left(\frac{3+\lambda_2}{2\lambda_1}\right)^2 = 4$$
$$\left(\frac{1-\lambda_2}{2\lambda_1}\right) - \left(\frac{3+\lambda_2}{2\lambda_1}\right) = 2.$$

From the second equation we have $\lambda_2 = -1 - 2\lambda_1$ and substituting into the first equation gives

$$\left(\frac{1+\lambda_1}{\lambda_1}\right)^2 + \left(\frac{1-\lambda_1}{\lambda_1}\right)^2 = 4,$$

which reduces to $\lambda_1^2 = 1$ or $\lambda_1 = \pm 1$. Since necessarily $\lambda_1 < 0$, we must take the case $\lambda_1 = -1$ giving $\lambda_2 = 1 > 0$ which is not allowed. Lastly consider the case $\lambda_1 < 0$ and $\lambda_2 = 0$; substitute into the constraint $x_1^2 + x_2^2 = 4$, which is tight because $z_1 = 0$, to get

$$\left(\frac{1}{2\lambda_1}\right)^2 + \left(\frac{3}{2\lambda_1}\right)^2 = 4,$$

and solve to obtain $\lambda_1 = -\sqrt{10}/4$. Thus the optimal values for the problem (and the Lagrange multipliers for the Lagrangian Sufficiency Theorem) are

$$x_1^* = -\sqrt{2/5}, \quad x_2^* = 3\sqrt{2/5}, \quad \lambda_1^* = -\sqrt{10}/4, \quad \lambda_2^* = 0.$$

You should draw a picture to see that the optimum lies on the first constraint and 'inside' the second. □

1.4 Shadow prices

The most important approach to understanding the role of Lagrange multipliers in the solution of optimization problems is to regard the problem

$$\mathrm{P}: \quad \text{minimize } f(\boldsymbol{x}) \quad \text{subject to} \quad g(\boldsymbol{x}) = \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

as one of a family of problems indexed by the right-hand side in the constraints, $\boldsymbol{b}$. To this end let $\phi(\boldsymbol{b})$ be the minimum value of the objective function f subject to the constraints

in P, so that $\phi(\boldsymbol{b}) = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in X,\ g(\boldsymbol{x}) = \boldsymbol{b}\}$. With this notation then there is an important relation between the Lagrange multipliers for the problem and ϕ in that

$$\lambda_i^* = \frac{\partial \phi}{\partial b_i}, \quad i = 1, \ldots, m.$$

It should be noted that this relation is also true if the functional constraints in P are formulated as inequality constraints, $g(\boldsymbol{x}) \leqslant \boldsymbol{b}$, and $\phi(\boldsymbol{b}) = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in X,\ g(\boldsymbol{x}) \leqslant \boldsymbol{b}\}$. We illustrate by considering a generalization of the problem in Example 1.3.

Example 1.4 Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1 - 3x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 \leqslant b_1 \\ & x_1 + x_2 \leqslant b_2, \end{aligned}$$

where here b_1 and b_2 are positive constants, which we will assume satisfy $b_2^2 > \frac{2}{5} b_1$ (see below). We may add slack variables and carry through the same arguments as before to rule out both $\lambda_1 < 0$ and $\lambda_2 < 0$; we would need

$$\left(\frac{1-\lambda_2}{2\lambda_1}\right)^2 + \left(\frac{3+\lambda_2}{2\lambda_1}\right)^2 = b_1 \quad \text{and} \quad \left(\frac{1-\lambda_2}{2\lambda_1}\right) - \left(\frac{3+\lambda_2}{2\lambda_1}\right) = b_2.$$

From the second equation we have $\lambda_2 = -1 - b_2\lambda_1$ and substituting into the first equation gives

$$\left(\frac{2+b_2\lambda_1}{2\lambda_1}\right)^2 + \left(\frac{2-b_2\lambda_1}{2\lambda_1}\right)^2 = b_1,$$

which reduces to $\lambda_1 = -2/\sqrt{2b_1 - b_2^2}$; if $2b_1 \leqslant b_2^2$ it is clear that there can be no solution with both $\lambda_1 < 0$ and $\lambda_2 < 0$ while if $2b_1 > b_2^2$ then take $\lambda_2 = -1 + 2b_2/\sqrt{2b_1 - b_2^2}$ which gives $\lambda_2 > 0$ if $5b_2^2 > 2b_1$. As before, we must have the case $\lambda_1 < 0$ and $\lambda_2 = 0$; substitute into the constraint $x_1^2 + x_2^2 = b_1$ to get

$$\left(\frac{1}{2\lambda_1}\right)^2 + \left(\frac{3}{2\lambda_1}\right)^2 = b_1,$$

from which we deduce that the optimal values for the problem are

$$x_1^* = -\sqrt{b_1/10}, \quad x_2^* = 3\sqrt{b_1/10}, \quad \lambda_1^* = -\sqrt{10/(4b_1)}, \quad \lambda_2^* = 0.$$

We then have $\phi(\boldsymbol{b}) = x_1^* - 3x_2^* = -\sqrt{10b_1}$ from which we can confirm that

$$\frac{\partial \phi}{\partial b_1} = -\sqrt{10/(4b_1)} = \lambda_1^*, \quad \text{and} \quad \frac{\partial \phi}{\partial b_2} = 0 = \lambda_2^*.$$

You should think about what happens when $b_2^2 < \frac{2}{5}b_1$; compute $\phi(\boldsymbol{b})$ and confirm that $\lambda_i^* = \partial\phi/\partial b_i$ in this case. □

Because of the relationship $\lambda_i^* = \partial\phi/\partial b_i$ the Lagrange multipliers are often known as **shadow prices** for the problem. The terminology makes most intuitive sense if we formulate the problem as a maximization as

$$\mathrm{P}': \quad \text{maximize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) \leqslant \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

and we think of some company with production processes $i = 1, \ldots, n$ which are operated at levels x_i to give a total profit $f(\boldsymbol{x})$. There are m different raw materials, $j = 1, \ldots, m$ such that if the production processes are at levels $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ then an amount $g_j(\boldsymbol{x})$ of raw material j is consumed where the available supplies are b_j, $1 \leqslant j \leqslant m$. The fact that we require $\boldsymbol{x} \in X$ may represent some technological constraints (incidental to the consumption of raw materials). Then the problem is one of choosing the production levels with the aim of maximizing the profit subject to using no more than the available supply of the raw materials. Then $\phi(\boldsymbol{b}) = \sup\{f(\boldsymbol{x}) : \boldsymbol{x} \in X,\ g(\boldsymbol{x}) \leqslant \boldsymbol{b}\}$ is the maximal profit.

If one considers now how much the company should be prepared to pay to secure Δb_j extra units of raw material j, then one can see that the marginal price, p_j, that it would be prepared to pay should be given by

$$p_j \Delta b_j = \frac{\partial \phi}{\partial b_j} \Delta b_j \quad \text{or} \quad p_j = \frac{\partial \phi}{\partial b_j} = \lambda_j^*.$$

1.5 The dual problem

A further important idea coming from the Lagrangian approach to constrained optimization is the notion of a related optimization problem, known as the **dual** problem, to the original problem

$$\mathrm{P}: \quad \text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) = \boldsymbol{b}, \quad \boldsymbol{x} \in X,$$

known as the **primal** problem. In a large class of cases the solutions to the primal and dual problems are very closely related and studying the dual problem gives insight into the solution of the primal problem and vice-versa. To formulate the dual, let

$$\Lambda = \left\{ \boldsymbol{\lambda} : \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}) > -\infty \right\}$$

be the set of Lagrange multipliers for which the minimum of the Lagrangian for P is finite. For each $\boldsymbol{\lambda} \in \Lambda$ let $h(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda})$. The dual problem is

$$\mathrm{D}: \quad \text{maximize } h(\boldsymbol{\lambda}) \qquad \text{subject to} \quad \boldsymbol{\lambda} \in \Lambda.$$

The function h is the objective function for the dual problem and the set Λ is the set of feasible solutions to the dual. We let

$$\psi(\boldsymbol{b}) = \sup\{h(\boldsymbol{\lambda}) : \ \boldsymbol{\lambda} \in \Lambda\},$$

represent the optimal value of the dual problem. An important relation between the two problems is contained in the following simple result.

Theorem 1.5 **Weak duality** *If $\boldsymbol{x}$ is any feasible solution for the primal problem* P *and $\boldsymbol{\lambda}$ is any feasible solution for the dual problem* D *then*

$$h(\boldsymbol{\lambda}) \leqslant f(\boldsymbol{x}).$$

In particular, it follows that $\psi(\boldsymbol{b}) \leqslant \phi(\boldsymbol{b})$.

Proof. For any feasible $\boldsymbol{x}$, $g(\boldsymbol{x}) = \boldsymbol{b}$ which implies that $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x})$, and so

$$h(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}' \in X} L(\boldsymbol{x}', \boldsymbol{\lambda}) \leqslant L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x});$$

maximizing over the left-hand side and minimizing over the right-hand side shows that $\psi(\boldsymbol{b}) \leqslant \phi(\boldsymbol{b})$. □

Example 1.6 *The dual problem*

Consider the problem of Example 1.2 but with a general right-hand side $b > 0$,

$$\begin{array}{ll} \text{minimize} & 2\ln(1/x_1) + 3\ln(1/x_2) + \ln(1/x_3) \\ \text{subject to} & 3x_1 + 2x_2 + x_3 = b, \quad x_1 > 0,\ x_2 > 0,\ x_3 > 0. \end{array}$$

The Lagrangian is now

$$L(\boldsymbol{x}, \lambda) = 2\ln(1/x_1) + 3\ln(1/x_2) + \ln(1/x_3) + \lambda(b - 3x_1 - 2x_2 - x_3),$$

which is minimized, as before, at $x_1^*(\lambda) = -2/(3\lambda)$, $x_2^*(\lambda) = -3/(2\lambda)$, $x_3^*(\lambda) = -1/\lambda$ (and the choice of λ to make these feasible for the problem is $\lambda^* = -6/b$); note that $\Lambda = \{\lambda : \ \lambda < 0\}$, and substituting these values gives

$$h(\lambda) = \inf_{x_i > 0} L(\boldsymbol{x}, \lambda) = \lambda b + 6\ln(-\lambda) + 2\ln(3/2) + 3\ln(2/3) + 6.$$

The dual problem is to maximize $h(\lambda)$ in $\lambda < 0$; we see that the maximizing value is $\lambda^* = -6/b$ and we can verify that in this case $\phi(b) = \psi(b)$. □

Remarks

1. The weak duality result shows that the value of the objective function at a feasible solution for the dual problem provides a lower bound for the objective function for the primal problem at any feasible solution for the primal problem; the largest such lower bound is the value of the dual problem.
2. The conclusion in this example, that the maximum value of the dual problem equals the minimum value of the primal problem, is true for any problem that can be solved using the Lagrangian method, that is for any problem for which there exists a $\boldsymbol{\lambda}^*$ for which

$$\inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}^*) = \inf_{\boldsymbol{x} \in X_{\boldsymbol{b}}} L(\boldsymbol{x}, \boldsymbol{\lambda}^*) = \inf_{\boldsymbol{x} \in X_{\boldsymbol{b}}} f(\boldsymbol{x}),$$

   where these terms are $> -\infty$; the general proof of this fact, which is known as **strong duality**, is beyond the scope of this course, but we will exploit it later for a particular class of problems, viz. linear programming.
3. The primal problem here is formulated as a minimization; if the primal problem is formulated as a maximization, then the dual objective function is obtained by maximizing the Lagrangian and the dual problem involves a minimization.

4. To get a feel for why the maximization and minimization are interchanged when going between the primal and dual problems, note that if we maximize the Lagrangian in the Lagrange multipliers we get, for $\boldsymbol{x} \in X$,

$$\sup_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \sup_{\boldsymbol{\lambda}} \left[f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top (\boldsymbol{b} - g(\boldsymbol{x})\right] = \begin{cases} f(\boldsymbol{x}) & \text{if } g(\boldsymbol{x}) = \boldsymbol{b}, \\ \infty & \text{otherwise;} \end{cases}$$

   this is because the Lagrange multipliers here can be either sign so that if any entry of $\boldsymbol{b} - g(\boldsymbol{x})$ is non-zero the corresponding entry of $\boldsymbol{\lambda}$ can be taken large (either positive or negative) to get ∞ as the supremum. Thus the primal and dual problems are

$$\text{Primal: } \min_{\boldsymbol{x} \in X} \left[\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right] \qquad \text{Dual: } \max_{\boldsymbol{\lambda}} \left[\min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right],$$

   and when the primal and dual problems have the same value it is the same as the statement that $\min_{\boldsymbol{x} \in X} \max_{\boldsymbol{\lambda}} L = \max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x} \in X} L$.

1.6 Lagrangian necessity

One may ask under what circumstances will the problem

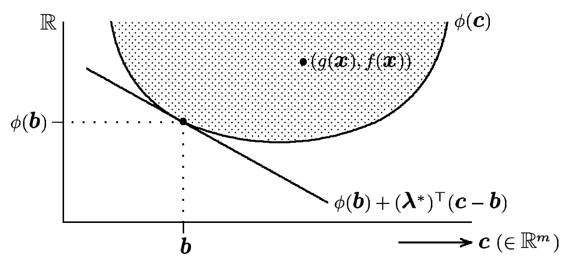
$$\text{P}: \quad \text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) = \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

(or its inequality version) be such that it can be guaranteed to be solved using the Lagrangian approach; that is, there will exist a $\boldsymbol{\lambda}^*$ satisfying the condition in Remark 2 of the previous section.

Consider the region $\{(g(\boldsymbol{x}), f(\boldsymbol{x})) : \ \boldsymbol{x} \in X\}$ in $\mathbb{R}^m \times \mathbb{R}$, illustrated in the diagram. The function $\phi(\cdot)$ is the lower boundary of that region, and whether or not the problem P can be solved by minimizing the Lagrangian for appropriate Lagrange multipliers depends on whether or not the function ϕ has a tangent plane, at the point $\boldsymbol{b}$, which lies entirely below the function (known as a **supporting hyperplane**).

Theorem 1.7 *For the problem* P, *there exist Lagrange multipliers* $\boldsymbol{\lambda}^*$ *satisfying*

$$\inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\lambda}^*) = \inf_{\boldsymbol{x} \in X_{\boldsymbol{b}}} L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$$



if and only if at the point $\boldsymbol{b}$ *the function* ϕ *has a supporting hyperplane with slope* $\boldsymbol{\lambda}^*$*; i.e.,*

$$\phi(\boldsymbol{c}) \geqslant \phi(\boldsymbol{b}) + (\boldsymbol{\lambda}^*)^\top (\boldsymbol{c}-\boldsymbol{b}) \quad \textit{for all } \boldsymbol{c}.$$

Proof. The condition (in the statement of the Theorem and in Remark 2) is equivalent to

$$\begin{aligned}\phi(\boldsymbol{b}) = \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{\lambda}^*) &= \inf_{\boldsymbol{x}\in X} \left[f(\boldsymbol{x}) + (\boldsymbol{\lambda}^*)^\top(\boldsymbol{b}-g(\boldsymbol{x})\right] \\ &= \inf_{\boldsymbol{c}} \inf_{\boldsymbol{x}\in X_{\boldsymbol{c}}} \left[f(\boldsymbol{x}) + (\boldsymbol{\lambda}^*)^\top(\boldsymbol{b}-g(\boldsymbol{x})\right] \\ &= \inf_{\boldsymbol{c}} \left[\phi(\boldsymbol{c}) + (\boldsymbol{\lambda}^*)^\top(\boldsymbol{b}-\boldsymbol{c})\right);\end{aligned}$$

it is immediate that this is equivalent to $\phi(\boldsymbol{b}) + (\boldsymbol{\lambda}^*)^\top(\boldsymbol{c}-\boldsymbol{b}) \leqslant \phi(\boldsymbol{c})$ for all $\boldsymbol{c}$. □

Remarks

1. There remains the question of how one can tell from properties of the functions f and g and of the set X whether or not the function ϕ has a supporting hyperplane at the relevant point. If ϕ has a supporting hyperplane at every point then necessarily it is a **convex** function. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$ and θ, $0 \leqslant \theta \leqslant 1$,

$$f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leqslant \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y});$$

   this is equivalent to requiring that the region $\{(y, f(\boldsymbol{x})) :\ y \geqslant f(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^n\}$ lying above the function f is a convex set in $\mathbb{R} \times \mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is convex if each component function is convex. Recall that a set $X \subseteq \mathbb{R}^n$ is convex if, for all $\boldsymbol{x}$, $\boldsymbol{y} \in X$ and θ, $0 \leqslant \theta \leqslant 1$, then $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \in X$. That is, if $\boldsymbol{x}$ and

$\boldsymbol{y}$ are in X then the whole line segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$ also lies in X. It follows (by induction on k) that if f is a convex function

$$f\left(\sum_{i=1}^{k} \theta_i \boldsymbol{x}_i\right) \leqslant \sum_{i=1}^{k} \theta_i f(\boldsymbol{x}_i),$$

   for any choice of $\boldsymbol{x}_i \in \mathbb{R}^n$, and $\theta_i \geqslant 0$, $i = 1, \ldots, k$, with $\sum_{i=1}^k \theta_i = 1$.

2. For the problem P, it is *sufficient* that f be a convex function, g be a linear function and X be a convex set to ensure that ϕ is convex and consequently the problem can be solved by minimizing the Lagrangian.
3. For the inequality formulation of the problem

$$\text{minimize } f(\boldsymbol{x}) \qquad \text{subject to} \quad g(\boldsymbol{x}) \leqslant \boldsymbol{b}, \quad \boldsymbol{x} \in X$$

   then ϕ is the lower boundary of the region $\{(g(\boldsymbol{x}) + \boldsymbol{z}, f(\boldsymbol{x})) :\ \boldsymbol{x} \in X,\ \boldsymbol{z} \geqslant 0\}$ in $\mathbb{R}^m \times \mathbb{R}$; in this case to ensure that ϕ is convex (and so as before, we can obtain the solution by minimizing the Lagrangian) it is *sufficient* that f and g are convex functions and X is a convex set.
4. It should be noted that the conditions in the situation of 3 are not enough in the context of 2. To see this, consider minimizing $f(x) = x^2$ subject to $g(x) = x^3 = b$, with $b > 0$, then $\phi(b) = b^{2/3}$ which is not convex.