5 Lagrangian Methods

5.1 The Lagrangian

Shortly, we will begin the course afresh in a rather more theoretical style. But first, we begin with some preliminary remarks, using problems P and D as examples. Recall that so far we have been more-or-less treating the relationships between P and D as interesting coincidences. In later lectures we establish more exactly the theoretical relationships between P and D. If you feel comfortable with the simplex algorithm, the following remarks may give a clue to what is going on, as well as introducing some of the theory from the next section. Skip this bit on first reading if it does not make sense. Recall

$$\begin{array}{lrcl} \text{P:} & \text{maximize} & x_1 + x_2 & \\ & \text{subject to} & x_1 + 2x_2 + z_1 & = & 6 \\ & & x_1 - x_2 + z_2 & = & 3 \\ & & x_1, x_2 & \geq & 0 \end{array}$$

The simplex algorithm, in proceeding to the optimum, just added multiples of constraints to the objective function, and stopped if it managed to get all coefficients ≤ 0. So, it was trying to find scalar multiples λ_1 and λ_2 of the two constraints so that

$$\max x_1 + x_2 - \lambda_1(x_1 + 2x_2 + z_1 - 6) - \lambda_2(x_1 - x_2 + z_2 - 3)$$

had a nice form. We write the **Lagrangian**

$$L(x, \lambda) = x_1 + x_2 - \lambda_1(x_1 + 2x_2 + z_1 - 6) - \lambda_2(x_1 - x_2 + z_2 - 3)$$

and rearranging terms

$$L(x, \lambda) = x_1(1 - \lambda_1 - \lambda_2) + x_2(1 - 2\lambda_1 + \lambda_2) - \lambda_1 z_1 - \lambda_2 z_2 + 6\lambda_1 + 3\lambda_2.$$

The λs are called **Lagrange multipliers**.

More specifically, the simplex algorithm was looking for a b.f.s. x, z to P and the multipliers λ_1, λ_2 such that the objective row was zero for basic variables and negative for non-basic variables.

$$\begin{array}{llll} x_1 = 0, \implies 1 - \lambda_1 - \lambda_2 < 0 & & x_2 = 0, \implies 1 - 2\lambda_1 + \lambda_2 < 0 \\ x_1 > 0, \implies 1 - \lambda_1 - \lambda_2 = 0 & & x_2 > 0, \implies 1 - 2\lambda_1 + \lambda_2 = 0 \end{array}$$

and

$$\begin{array}{llll} z_1 = 0, \implies \lambda_1 > 0 & & z_2 = 0, \implies \lambda_2 > 0 \\ z_1 > 0, \implies \lambda_1 = 0 & & z_2 > 0, \implies \lambda_2 = 0 \end{array}$$

But these conditions can be summarised as dual feasibility:

$$-v_1 = 1 - \lambda_1 - \lambda_2 \leq 0, \quad -v_2 = 1 - 2\lambda_1 + \lambda_2 \leq 0, \quad \lambda_1, \lambda_2 \geq 0,$$

and complementary slackness:

$$x_1(1 - \lambda_1 - \lambda_2) = 0, \quad x_2(1 - 2\lambda_1 + \lambda_2) = 0, \quad \lambda_1 z_1 = 0, \quad \lambda_2 z_2 = 0.$$

If all these are satisfied, then objective $= \lambda^\top b = c^\top x$. See Theorem 2.4.

This is where the dual problem comes from. Lagrange multipliers are the same thing as dual variables (and shadow prices, and two more names we will encounter before the course ends.)

5.2 The Lagrangian sufficiency theorem

Given a general optimization under constraints problem,

$$\text{P:} \quad \min\ f(x) \quad \text{s.t. } g(x) = b, \quad x \in X,$$

with $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ (n variables and m constraints), the Lagrangian is

$$L(x, \lambda) = f(x) - \lambda^\top (g(x) - b),$$

with $\lambda \in \mathbb{R}^m$ (one for each constraint). The Lagrangian is a function $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

(Some motivation for this definition can be found at the end of the last section, but note that we do not require anything to be linear for this definition, or the results that follow.) The following theorem will be useful, and is easy to prove.

Theorem 5.1 (Lagrangian sufficiency theorem) *If x^* and λ^* exist such that x^* is feasible for P and*

$$L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \forall x \in X,$$

then x^* is optimal for P.

Proof. Define

$$X_b = \{x : x \in X \text{ and } g(x) = b\}.$$

Note that $X_b \subseteq X$ and that for any $x \in X_b$

$$L(x, \lambda) = f(x) - \lambda^\top (g(x) - b) = f(x).$$

Now

$$f(x^*) = L(x^*, \lambda^*) \leq L(x, \lambda^*) = f(x), \quad \forall x \in X_b.$$

The inequality in the above holds for all $x \in X$ by assumption, and so in particular for $x \in X_b$. Thus x^* is optimal for P. ■

Remarks.

1. Note the 'If' which starts the statement of the theorem. There is no guarantee that we can find a λ^* satisfying the conditions of the theorem for general problems P. (However, we will see that there is a large class of problems for which λ^* do exist.)

2. At first sight the theorem offers us a method for testing that a solution x^* is optimal for P without helping us to find x^* if we don't already know it. Certainly we will sometimes use the theorem this way. But for some problems, there is a way we can use the theorem to find the optimal x^*.

5.3 Strategy to solve problems with the Lagrangian sufficiency theorem

Attempt to find x^*, λ^* satisfying the conditions of the theorem as follows.

1. For each λ solve the problem

$$\text{minimize } L(x, \lambda) \text{ subject to } x \in X.$$

   Note that the only constraints involved in this problem are $x \in X$ so this should be an easier problem to solve than P.

2. Define the set

$$Y = \{\lambda : \min_{x \in X} L(x, \lambda) > -\infty\}.$$

   If we obtain $-\infty$ for the minimum in step 1 then that λ is no good. We consider only those $\lambda \in Y$ for which we obtain a finite minimum.

3. For $\lambda \in Y$, the minimum will be obtained at some $x(\lambda)$ (that depends on λ in general). Typically, $x(\lambda)$ will not be feasible for P.

4. Adjust $\lambda \in Y$ so that $x(\lambda)$ is feasible. If $\lambda^* \in Y$ exists such that $x^* = x(\lambda^*)$ is feasible then x^* is optimal for P by the theorem.

5.4 Example: use of the Lagrangian sufficiency theorem

$$\begin{aligned} &\text{minimize } x_1 - x_2 - 2x_3 \\ &\text{s.t. } x_1 + x_2 + x_3 = 5 \\ &x_1^2 + x_2^2 = 4 \\ &x \in X = \mathbb{R}^3. \end{aligned}$$

Solution. Since we have two constraints we take $\lambda \in \mathbb{R}^2$ and write the Lagrangian

$$\begin{aligned} L(x, \lambda) &= x_1 - x_2 - 2x_3 - \lambda_1(x_1 + x_2 + x_3 - 5) - \lambda_2(x_1^2 + x_2^2 - 4) \\ &= \left[x_1(1 - \lambda_1) - \lambda_2 x_1^2\right] + \left[x_2(-1 - \lambda_1) - \lambda_2 x_2^2\right] \\ &\quad + \left[x_3(-2 - \lambda_1)\right] + 5\lambda_1 + 4\lambda_2. \end{aligned}$$

We first try to minimize $L(x, \lambda)$ for fixed λ in $x \in \mathbb{R}^3$. Notice that we can minimize each square bracket separately.

First notice that $-x_3(2 + \lambda_1)$ has minimum $-\infty$ unless $\lambda_1 = -2$. So we only want to consider $\lambda_1 = -2$.

Now observe that the terms in x_1, x_2 have a finite minimum only if $\lambda_2 < 0$ in which case the minimum occurs at

$$\begin{aligned} 1 - \lambda_1 - 2\lambda_2 x_1 = 0 &\implies x_1 = 3/2\lambda_2 \\ -1 - \lambda_1 - 2\lambda_2 x_2 = 0 &\implies x_2 = 1/2\lambda_2. \end{aligned}$$

So Y is the set

$$Y = \{\lambda : \lambda_1 = -2,\ \lambda_2 < 0\},$$

and for $\lambda \in Y$ the minimum of $L(x, \lambda)$ occurs at $x(\lambda) = (3/2\lambda_2, 1/2\lambda_2, x_3)^\top$.

Now to find a feasible $x(\lambda)$ we need

$$x_1^2 + x_2^2 = 4 \implies \frac{9}{4\lambda_2^2} + \frac{1}{4\lambda_2^2} = 4 \implies \lambda_2 = -\sqrt{5/8}.$$

So $x_1 = -3\sqrt{2/5}$, $x_2 = -\sqrt{2/5}$ and $x_3 = 5 - x_1 - x_2 = 5 + 4\sqrt{2/5}$.

The conditions of the Lagrangian sufficiency theorem are satisfied by

$$x^* = \left(-3\sqrt{2/5}, -\sqrt{2/5}, 5 + 4\sqrt{2/5}\right)^\top \text{ and } \lambda^* = \left(-2, -\sqrt{5/8}\right)^\top.$$

So x^* is optimal. ■