Appendix A

Lagrangian Methods for Constrained Optimization

A.1 Regional and functional constraints

Throughout this book we have considered optimization problems that were subject to constraints. These include the problem of allocating a finite amounts of bandwidth to maximize total user benefit (page 17), the social welfare maximization problem (page 129) and the time of day pricing problem (page 213). We make frequent use of the Lagrangian method to solve these problems. This appendix provides a tutorial on the method. Take, for example,

$$\texttt{NETWORK} : \underset{x\geq 0}{\text{maximize}} \sum_{r=1}^{n_r} w_r \log x_r \,, \text{ subject to } Ax \leq C \,,$$

posed on page 271. This is an example of the generic constrained optimization problem:

$$P : \ \underset{x\in X}{\text{maximize}}\, f(x), \quad \text{subject to } g(x) = b \,.$$

Here f is to be maximized subject to constraints that are of two types. The constraint $x \in X$ is a **regional constraint**. For example, it might be $x \geq 0$. The constraint $g(x) = b$ is a **functional constraint**. Sometimes the functional constraint is an inequality constraint, like $g(x) \leq b$. But if it is, we can always add a **slack variable**, z, and re-write it as the equality constraint $g(x) + z = b$, re-defining the regional constraint as $x \in X$ and $z \geq 0$. To illustrate things we shall use the `NETWORK` problem with just one resource constraint

$$P_1 : \underset{x\geq 0}{\text{maximize}} \sum_{i=1}^{n} w_i \log x_i, \quad \text{subject to } \sum_{i=1}^{n} x_i = b \,,$$

where b is a positive number.

A.2 The Lagrangian method

The solution of a constrained optimization problem can often be found by using the so-called **Lagrangian method**. We define the **Lagrangian** as

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

For P_1 it is

$$L_1(x,\lambda) = \sum_{i=1}^{n} w_i \log x_i + \lambda \left(b - \sum_{i=1}^{n} x_i \right) .$$

In general, the Lagrangian is the sum of the original objective function and a term that involves the functional constraint and a 'Lagrange multiplier' λ. Suppose we ignore the functional constraint and consider the problem of maximizing the Lagrangian, subject only to the regional constraint. This is often an easier problem than the original one. The value of x that maximizes $L(x,\lambda)$ depends on the value of λ. Let us denote this optimizing value of x by $x(\lambda)$.

For example, since $L_1(x,\lambda)$ is a concave function of x it has a unique maximum at a point where f is stationary with respect to changes in x, i.e., where

$$\partial L_1/\partial x_i = w_i/x_i - \lambda = 0 \quad \text{for all } i\,.$$

Thus $x_i(\lambda) = w_i/\lambda$. Note that $x_i(\lambda) > 0$ for $\lambda > 0$, and so the solution lies in the interior of the feasible set.

Think of λ as knob that we can turn to adjust the value of x. Imagine turning this knob until we find a value of λ, say $\lambda = \lambda^*$, such that the functional constraint is satisfied, i.e., $g(x(\lambda^*)) = b$. Let $x^* = x(\lambda^*)$. Our claim is that x^* solves P. This is the so-called Lagrangian Sufficiency Theorem, which we state and prove shortly. First note that, in our example, $g(x(\lambda)) = \sum_i w_i/\lambda$. Thus choosing $\lambda^* = \sum_i w_i/b$, we have $g(x(\lambda^*)) = b$. The next theorem shows that $x = x(\lambda^*) = w_i b/\sum_j w_j$ is optimal for P_1.

Theorem 5 (Lagrangian Sufficiency Theorem) *Suppose there exist $x^* \in X$ and λ^*, such that x^* maximizes $L(x,\lambda^*)$ over all $x \in X$, and $g(x^*) = b$. Then x^* solves P.*

Proof.

$$\begin{aligned}
\max_{\substack{x\in X\\ g(x)=b}} f(x) &= \max_{\substack{x\in X\\ g(x)=b}} [f(x) + \lambda^*(b-g(x))] \\
&\le \max_{x\in X} [f(x) + \lambda^*(b-g(x))] \\
&= f(x^*) + \lambda^*(b-g(x^*))] \\
&= f(x^*)
\end{aligned}$$

Equality in the first line holds because we have simply added 0 on the right hand side. The inequality in the second line holds because we have enlarged the set over which maximization takes place. In the third line we use the fact that x^* maximizes $L(x,\lambda^*)$ and in the fourth line we use $g(x^*) = b$. But x^* is feasible for P, in that it satisfies the regional and functional constraints. Hence x^* is optimal. ■

Multiple constraints

If g and b are vectors, so that $g(x) = b$ expresses more than one constraint, then we would write

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

where the vector λ now has one component for each constraint. For example, the Lagrangian for `NETWORK` is

$$L(x,\lambda) = \sum_{r=1}^{n_r} w_r \log x_r + \sum_j \lambda_j (C_j - \sum_j A_{jr} x_r - z_j)$$

where z_j is the slack variable for the jth constraint.

A.3 When does the method work?

The Lagrangian method is based on a 'sufficiency theorem'. The means that method can work, but need not work. Our approach is to write down the Lagrangian, maximize it, and then see if we can choose λ and a maximizing x so that the conditions of the Lagrangian Sufficiency Theorem are satisfied. If this works, then we are happy. If it does not work, then too bad. We must try to solve our problem some other way. The method worked for P_1 because we could find an appropriate λ^*. To see that this is so, note that as λ increases from 0 to ∞, $g(x(\lambda))$ decreases from ∞ to 0. Moreover, $g(x(\lambda))$ is continuous in λ. Therefore, given positive b, there must exist a λ for which $g(x(\lambda)) = b$. For this value of λ, which we denote λ^*, and for $x^* = x(\lambda^*)$ the conditions of the Lagrangian Sufficiency Theorem are satisfied.

To see that the Lagrangian method does not always work, consider the problem

$$P_2 : \text{ minimize} -x, \text{ subject to } x \geq 0 \text{ and } \sqrt{x} = 2\,.$$

This cannot be solved by the Lagrangian method. If we minimize

$$L(x, \lambda) = -x + \lambda(2 - \sqrt{x})$$

over $x \geq 0$, we get a minimum value of $-\infty$, no matter what we take as the value of λ. This is clearly not the right answer. So the Lagrangian method fails to work. However, the method does work for

$$P_2' : \text{minimize} -x, \text{ subject to } x \geq 0 \text{ and } x^2 = 16$$

Now

$$L(x, \lambda) = -x + \lambda(16 - x^2)$$

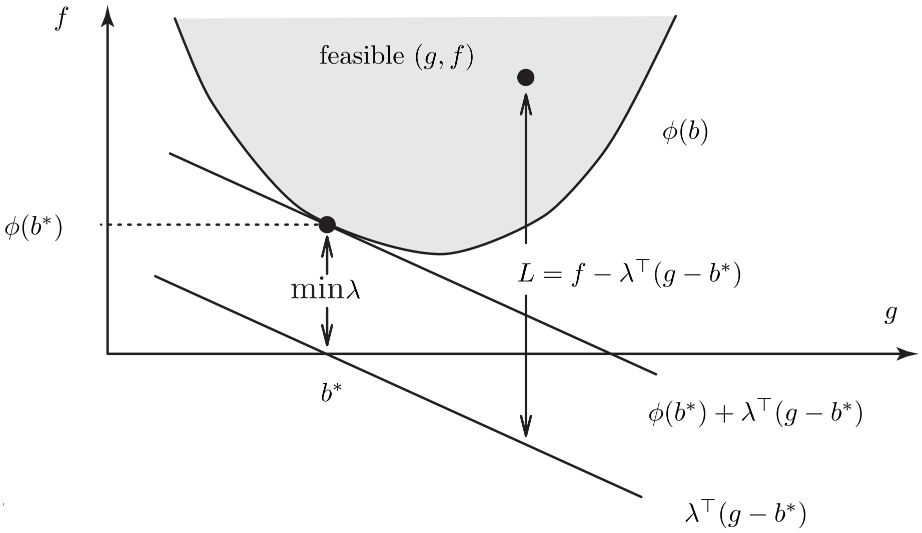
If we take $\lambda^* = -1/8$ then $\partial L/\partial x = -1 + x/4$ and so $x^* = 4$. Note that P_2 and P_2' are really the same problem, except that the functional constraint is expressed differently. Thus whether or not the Lagrangian method will work can depend upon how we formulate the problem.

We can say something more about when the Lagrangian method will work. Let $P(b)$ be the problem: *minimize* $f(x)$*, such that* $x \in X$ *and* $g(x) = b$. Define $\phi(b)$ as $\min f(x)$, subject to $x \in X$ and $g(x) = b$. Then the Lagrangian method works for $P(b^*)$ if and only if there is a line that is tangent to $\phi(b)$ at b^* and lies completely below $\phi(b)$. This happens if $\phi(b)$ is a convex function of b, but this is a difficult condition to check. A set of sufficient conditions that are easier to check are provided in the following theorem. These conditions do not hold in P_2 as $g(x) = \sqrt{x}$ is not a convex function of x. In P_2' the sufficient conditions are met.

Theorem 6 *If f and g are convex functions, X is a convex set, and x^* is an optimal solution to P, then there exist Lagrange multipliers $\lambda \in \mathbb{R}^m$ such that $L(x^*, \lambda) \leq L(x, \lambda)$ for all $x \in X$.*

Remark. Recall that f is a **convex function** if for all $x_1, x_2 \in X$ and $\theta \in [0, 1]$, we have $f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2)$. X is a **convex set** if for all $x_1, x_2 \in X$ and $\theta \in [0, 1]$, we have $\theta x_1 + (1 - \theta)x_2 \in X$. Furthermore, f is a **concave function** if $-f$ is convex.

The proof of the theorem proceeds by showing that $\phi(b)$ is convex. This implies that for each b^* there is a tangent hyperplane to $\phi(b)$ at b^* with the graph of $\phi(b)$ lying entirely above it. This uses the so-called 'supporting hyperplane theorem' for convex sets, which is geometrically obvious, but some work to prove. Note the equation of the hyperplane will be $y = \phi(b^*) + \lambda^\top(b - b^*)$ for some multipliers λ. This λ can be shown to be the required vector of Lagrange multipliers and the picture below gives some geometric intuition as to why the Lagrange multipliers λ exist and why these λs give the rate of change of the optimum $\phi(b)$ with b.



A.4 Shadow prices

The maximizing x, the appropriate value of λ and maximal value of f all depend on b. What happens if b changes by a small amount? Let the maximizing x be $x^*(b)$. Suppose the Lagrangian method works and let $\lambda^*(b)$ denote the appropriate value of λ. As above, let $\phi(b)$ denote the maximal value of f. We have

$$\phi(b) = f(x^*(b)) + \lambda^* \left[b - g(x^*(b)\right] .$$

So simply differentiating with respect to b, we have

$$\begin{aligned}\frac{\partial}{\partial b}\phi(b) &= \sum_{i=1}^{n} \frac{\partial x_i^*}{\partial b}\frac{\partial}{\partial x_i^*}\left\{f(x^*(b)) + \lambda^*[b - g(x^*(b)]\right\} \\ &+ \frac{\partial}{\partial b}\{f(x^*(b)) + \lambda^*[b - g(\bar{x}(b)]\} + \frac{\partial \lambda^*}{\partial b}\frac{\partial}{\partial \lambda^*}\{f(x^*(b)) + \lambda^*[b - g(x^*(b)]\} \\ &= 0 + \lambda^* + [b - g(x^*(b)]\frac{\partial \lambda^*}{\partial b} ,\end{aligned}$$

where the first term on the right hand side is 0 because $L(x, \lambda^*)$ is stationary with respect to x_i at $x = x^*$ and the third term is zero because $b - g(x^*(b)) = 0$. Thus

$$\frac{\partial}{\partial b}\phi(b) = \lambda^*$$

and λ can be interpreted as the rate at which the maximized value of f increases with b, for small increases around b. For this reason, the Lagrange multiplier λ^* is also called a

shadow price, the idea being that if b increases to $b+\delta$ then we should be prepared to pay $\lambda^*\delta$ for the increase we receive in f.

It can happen that at the optimum none, one, or several constraints are active. E.g., with constraints $g_1(x) \le b_1$ and $g_2(x) \le b_2$ it can happen that at the optimum $g_1(x^*) = b_1$ and $g_2(x^*) < b_2$. In this case we will find have $\lambda_2^* = 0$. This makes sense. The second constraint is not limiting the maximization of f and so the shadow price of b_2 is zero.

A.5 The dual problem

By similar reasoning to that we used in the proof of the Lagrangian sufficiency theorem, we have that for any λ

$$\begin{aligned}\phi(b) &= \max_{\substack{x\in X\\ g(x)=b}} f(x)\\ &= \max_{\substack{x\in X\\ g(x)=b}} [f(x)+\lambda(b-g(x))]\\ &\le \max_{x\in X}[f(x)+\lambda(b-g(x))]\,.\end{aligned}$$

The right hand side provides an upper bound on $\phi(b)$. We make this upper bound as tight as possible by minimizing over λ, so that we have

$$\phi(b) \le \min_{\lambda}\max_{x\in X}[f(x)+\lambda^*(b-g(x))]\,.$$

The right hand side above defines an optimization problem, called the **dual problem**. The original problem is called the **primal problem**. If the primal can be solved by the Lagrangian method then the inequality above is an equality and the solution to the dual problem is just $\lambda^*(b)$. If the primal cannot be solved by the Lagrangian method we will have a strict inequality, the so-called **duality gap**.

The dual problem is interesting because it can sometimes be easier to solve, or because it formulates the problem in an illuminating way. The dual of P_1 is

$$\operatorname*{minimize}_{\lambda} \sum_{i=1}^{n} w_i \log(w_i/\lambda) + \lambda\left(b - \sum_{i=1}^{n} w_i/\lambda\right)\,,$$

where we have inserted $x_i = w_i/\lambda$, after carrying out the inner maximization over x. This is a convex function of λ. Differentiating with respect to λ, one can check that the stationary point is the maximum, and $\sum_i w_i/\lambda = b$. This gives λ, and finally, as before

$$\phi(b) = \sum_{i=1}^{n} w_i \log\left(\frac{w_i}{\sum_j w_j} b\right)\,.$$

The dual plays a particularly important role in the theory of **linear programming**. A linear program, such

$$P:\ \text{maximize}\, c^\top x,\ \text{subject to } x\ge 0 \text{ and } Ax\le b$$

is one in which both the objective function and constraints are linear. Here $x, c\in\mathbb{R}^n$, $b\in\mathbb{R}^m$ and A is a $m\times n$ matrix. The dual of P is D.

$$D:\ \text{minimize}\, \lambda^\top b,\ \text{subject to } \lambda\ge 0 \text{ and } y^\top A\ge c^\top$$

D is another linear program and the dual of D is P. The decision variables in D, i.e., $\lambda_1,\ldots,\lambda_m$, are the Lagrange multipliers for the m constraints expressed by $Ax\le b$ in P.

Reference

For further details on Lagrangian methods of constrained optimization, see the course notes of Weber (1998).