

Operations Research Techniques and Algorithms (620-361)

Dr Yao-ban Chan

y.chan@ms.unimelb.edu.au

Telephone: 8344 9073

Office: Room 198, Richard Berry Building

Christina Burt

c.burt@ms.unimelb.edu.au

Telephone: 8344 1797

Office: 139 Barry St

April 16, 2008

Today's Lecture

Constrained optimisation

Second order theorems and the Lagrange multipliers

We will now prove the second-order theorems formally.

For Theorem 5, first note that since f is convex, any local minimum is also a global minimum.

Now, applying proof by contradiction, assume that f is convex and h is affine and the Lagrangian conditions hold at x^* , but x^* is not a minimum.

Then there must exist another feasible point x' for which $f(x') < f(x^*)$.

Because h is affine, and both x' and x^* are feasible, the line joining x' to x^* must also be feasible.

By the convexity of f , it must lie below the line joining $(x', f(x'))$ and $(x^*, f(x^*))$. In particular, if we set $d = x' - x^*$, this means that the directional derivative of f at x^* in the direction of d must be less than or equal to

$$\frac{f(x') - f(x^*)}{x' - x^*} < 0.$$

But from the Lagrangian conditions, we know that

$$\nabla f(x^*) + \nabla h(x^*)\eta^* = 0$$

and therefore the directional derivative of f at x^* in the direction of d is

$$\begin{aligned}\langle \nabla f(x^*), d \rangle &= \nabla f(x^*)^T d \\ &= -[\nabla h(x^*)\eta^*]^T d \\ &= -(\eta^*)^T (\nabla h(x^*)^T d).\end{aligned}$$

Given that $d = x' - x^*$, and $h(x') = h(x^*) = 0$, we can say that d is parallel to each of the level surfaces $h_j(x) = 0$. But $\nabla h_j(x^*)$ is normal to these level surfaces at x^* , which means that $\nabla h_j(x^*)^T d = 0$, which in turn implies

$$\nabla h(x^*)^T d = 0.$$

From our calculations, this shows that the directional derivative of f at x^* in the direction of d must be 0, which contradicts the fact that it is strictly less than 0. Such a contradiction can only imply that there is no such x' , i.e. that x^* is indeed a local (and global) minimum.

Now we prove Theorem 6.

Let d be a feasible direction at x^* , i.e. $d \in \mathcal{C}(x^*)$. Using the Taylor series expansion, for small t ,

$$\begin{aligned} f(x^* + td) &= f(x^*) + \langle \nabla f(x^*), td \rangle + \frac{1}{2} \langle td, \nabla^2 f(x^*)(td) \rangle + o(t^2) \\ &= f(x^*) + t \nabla f(x^*)^T d + \frac{1}{2} t^2 d^T \nabla^2 f(x^*) d + o(t^2). \end{aligned}$$

As in the argument above, we know that d is parallel to each of the level surfaces $h_j(x) = 0$ at x^* . This means that

$$\nabla h(x^*)^T d = 0.$$

Since the Lagrangian conditions are satisfied, we can use the same argument:

$$\begin{aligned}t \nabla f(x^*)^T d &= -t [\nabla h(x^*) \eta^*]^T d \\ &= -t (\eta^*)^T h(x^*)^T d \\ &= 0.\end{aligned}$$

We also assume that the Hessian of the Lagrangian is positive definite on $\mathcal{C}(x^*)$. This means that for all feasible directions d ,

$$d^T \left[\nabla^2 f(x^*) + \sum_{j=1}^q \eta_j^* \nabla^2 h_j(x^*) \right] d > 0.$$

It can be shown (but would take too long to do it here) that in fact

$$d^T \nabla^2 f(x^*) d > 0.$$

Since

$$f(x^* + td) = f(x^*) + \frac{1}{2} t^2 d^T \nabla^2 f(x^*) d + o(t^2),$$

and for small t the second term will always dominate the third, this means that as we travel along any feasible direction, the function will increase. Therefore x^* is a local minimum and the theorem is (sort of) proved.

What do Lagrange multipliers mean?

It is not immediately obvious what Lagrange multipliers mean. Each optimal point has one Lagrange multiplier for each constraint, but how can we interpret their values?

Lagrange multipliers can be interpreted as *shadow prices* (economically), or mathematically as rates of change of the optimal f as the level of constraint changes.

To see this, consider each constraint as a resource constraint that we must satisfy.

For example, for the problem

$$\begin{aligned} \min \quad & f(x) = x_1^2 + \frac{1}{4}x_2^2 \\ \text{s.t.} \quad & h(x) = x_1 + x_2 + 1 = 0 \end{aligned}$$

we have a minimum at $(x_1, x_2) = (-\frac{1}{5}, -\frac{4}{5})$, with Lagrangian multiplier $\eta^* = \frac{2}{5}$.

We can think of the constraint as representing some kind of resource, of which we have -1 in total (it is obviously a strange kind of resource!), and of which we use $x_1 + x_2$ units.

Now, if we increase the level of resource that we have available to use by a small amount, so the constraint becomes

$$x_1 + x_2 = -1 + \Delta,$$

then we end up with a new optimal solution. This optimal solution is probably only slightly different from the old one, so we say it is $x^* + \Delta x$, where $\Delta x = (\Delta x_1, \Delta x_2)$.

Plugging this value into the new constraint shows us that

$$x_1^* + \Delta x_1 + x_2^* + \Delta x_2 = -1 + \Delta$$

and since x^* is feasible for the original problem, this means that

$$\Delta x_1 + \Delta x_2 = \Delta.$$

Then the new optimal f is given by the first-order Taylor series expansion

$$f(x^* + \Delta x) \approx f(x^*) + \nabla f(x^*)^T \Delta x.$$

Given that $\nabla f(x^*) = (-\frac{2}{5}, -\frac{2}{5}) = -\eta^*(1, 1)$, this shows us that

$$\begin{aligned} f(x^* + \Delta x) - f(x^*) &\approx -\eta^*(1, 1) \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} \\ &= -\eta^*(\Delta x_1 + \Delta x_2) \\ &= -\eta^* \Delta. \end{aligned}$$

So if we change our resource value by a small amount, the value that our optimal function value changes by is proportional to that small amount, and the proportionality constant is $-\eta^*$.

Another way of putting this is

$$\frac{f(x^* + \Delta x) - f(x^*)}{\Delta} \approx -\eta^*.$$

So as we change the amount of resource we have available, the rate of change of our optimal function value is $-\eta^*$.

We call this a *shadow price* because it is the maximum amount we would pay to have another unit of resource available to us (if we have to pay exactly that amount, the cost and the extra 'revenue' balance out).

In a more general case, suppose that we have to minimise a function $f(x)$ subject to a single linear equality constraint $a^T x = b$. Suppose further that (x^*, η^*) is a local minimum and corresponding Lagrange multiplier for this problem. By Lagrange's condition,

$$\nabla f(x^*) = -\eta^* \nabla h(x^*) = -\eta^* a.$$

Now suppose that the right-hand side of the constraint is changes by an amount Δ to $b + \Delta$, and this changes the local minimum to a point $x^* + \Delta x$. Then

$$b + \Delta = a^T (x^* + \Delta x) = a^T x^* + a^T \Delta x.$$

This implies that

$$\Delta = a^T \Delta x.$$

The amount that the optimal f value is changed because of the change in the constraint is

$$\begin{aligned} f(x^* + \Delta x) - f(x^*) &\approx \nabla f(x^*)^T \Delta x \\ &= -\eta^* a^T \Delta x \\ &= -\eta^* \Delta. \end{aligned}$$

So the same conclusions can be drawn: a small amount of change in the right-hand side of a constraint results in a proportional change in the optimal function value, and the proportionality constant is $-\eta^*$.

Both these cases are specialised cases of the general theorem, stated below.

Theorem 7: *Consider the family of equality-constrained nonlinear programs*

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = u \end{array}$$

where $u \in \mathbb{R}^q$. Then there exists an open sphere S , centred at the origin, such that for every $u \in S$, there is an $x(u)$ and $\eta(u)$ which are the local minimum and Lagrange multipliers for the NLP. Furthermore, $x(u)$ and $\eta(u)$ are C^1 , and for all $u \in S$,

$$\nabla_u f(x(u)) = -\eta(u).$$

In other words, $-\eta(u)$ is the rate at which the optimal value of f changes, as u is changed.

Example: A cylinder has length l and radius r . If we want to make a cylindrical tin can that holds a unit amount of volume, using the least amount of material, what are the optimal values for l and r ?