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

May 9, 2008

We illustrate this result with an example. Consider the nonlinear program

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.t.} & -x_1 - x_2 + 4 \leq 0 \\ & x_1, x_2 \geq 0. \end{array}$$

It can be seen by inspection that this NLP has a unique KKT point $x^* = (2, 2)$ with multipliers $\lambda^* = (4, 0, 0)$.

Now

$$\begin{aligned}\phi(x) &= L(x, \lambda^*) \\ &= x_1^2 + x_2^2 + 4(-x_1 - x_2 + 4) + 0(-x_1) + 0(-x_2) \\ &= x_1^2 + x_2^2 - 4x_1 - 4x_2 + 16.\end{aligned}$$

To find the minimum of $\phi(x)$ over all x, we simply take $\nabla\phi(x) = 0$.

$$\nabla \phi(x) = (2x_1 - 4, 2x_2 - 4) = (0, 0).$$

This is solved by $x = (2, 2)$. Checking that it is in fact a minimum:

$$\nabla^2 \phi(x) = \left[\begin{array}{cc} 2 & 0 \\ 0 & 2 \end{array} \right]$$

which is obviously positive definite. Therefore $(2, 2)$ minimises $L(x, \lambda^*)$, and we see that it is indeed equal to x^*.

Given the previous result, we can ask: if we want to convert a constrained problem to an unconstrained problem, why can't we just minimise the Lagrangian as an unconstrained problem?

Given the previous result, we can ask: if we want to convert a constrained problem to an unconstrained problem, why can't we just minimise the Lagrangian as an unconstrained problem?

We could — if we knew the optimal KKT multipliers! But we usually don't.

We have seen that given specific KKT multipliers λ^* and η^*, we can consider the Lagrangian as a function of x alone. In the next theorem, we also do the opposite: given a specific point x^*, we can consider the Lagrangian as a function of λ and η!

Theorem. The triple (x^*, λ^*, η^*) is a KKT point of a convex program if and only if, for all $\lambda \geq 0$ and all x and η,

$$L(x^*, \lambda, \eta) \leq L(x^*, \lambda^*, \eta^*) \leq L(x, \lambda^*, \eta^*).$$

This inequality is known as the Saddle Inequality and the theorem as the saddlepoint theorem.

Note that this result goes two ways: previously, we knew that if (x^*, λ^*, η^*) is a KKT point, then it minimises the Lagrangian with λ^* and η^* fixed. However, the theorem also tells us that if x^* minimises the Lagrangian, it is also a KKT point!

Proof. First we consider the forward implication: if (x^*, λ^*, η^*) is a KKT point, then the Saddle Inequality holds. The second inequality is just the Corollary above, so we already know it to be true.

Since (x^*, λ^*, η^*) is a KKT point, from KKTb and KKTc we know that

$$(\lambda^*)^T g(x^*) = 0$$

and

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

So

$$L(x^*, \lambda^*, \eta^*) = f(x^*) + (\lambda^*)^T g(x^*) + (\eta^*)^T h(x^*) = f(x^*).$$

Now consider any $\lambda \geq 0$ and any η. Since $g(x^*) \leq 0$ by KKTb, we have

$$\lambda^T g(x^*) \leq 0$$

and

$$\eta^T h(x^*) = 0$$

from KKTc.

Therefore

$$
\begin{aligned}
L(x^*, \lambda, \eta) &= f(x^*) + \lambda^T g(x^*) + \eta^T h(x^*) \\
&\leq f(x^*) \\
&= L(x^*, \lambda^*, \eta^*).
\end{aligned}
$$

This is the left inequality of the Saddle Inequality.

Next we consider the backward implication: if the Saddle Inequality holds then (x^*, λ^*, η^*) is a KKT point.

The right-hand side of the Saddle Inequality says that x^* minimises $L(x, \lambda^*, \eta^*)$ over all x. Therefore, x^* must be a stationary point of $L(x, \lambda^*, \eta^*)$, considered as a function of x, and therefore

$$\nabla_x L(x, \lambda^*, \eta^*) = 0.$$

This is KKTa.

The left-hand side of the Saddle Inequality says that (λ^*, η^*) maximises $L(x^*, \lambda, \eta)$ (considered as a function of (λ, η)) over all $\lambda \geq 0$ and all η. If you consider the KKT multipliers as variables themselves, we see that this is another optimisation problem:

$$\begin{array}{ll} \min & -L(x^*, \lambda, \eta) \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

We denote this problem by (P).

This problem must also have a Lagrangian function, which we denote by $L^P(\lambda, \eta, \alpha)$, where α is the KKT multipliers:

$$\begin{aligned} L^P(\lambda, \eta, \alpha) &= -L(x^*, \lambda, \eta) + \alpha^T(-\lambda) \\ &= -L(x^*, \lambda, \eta) - \alpha^T \lambda. \end{aligned}$$

Since (λ^*, η^*) solves (P), there must exist KKT multipliers α^* such that the KKT conditions for (P) hold for the triple $(\lambda^*, \eta^*, \alpha^*)$. For KKTa, this turns out to be

$$\nabla_\lambda L^P(\lambda^*, \eta^*, \alpha^*) = 0$$

and

$$\nabla_\eta L^P(\lambda^*, \eta^*, \alpha^*) = 0.$$

For KKTb, this turns out to be

$$-\lambda^* \leq 0, \alpha^* \geq 0, (\alpha^*)^T \lambda^* = 0.$$

Now from the KKTa condition of (P),

$$\begin{aligned} \nabla_\lambda L^P(\lambda^*, \eta^*, \alpha^*) &= -\nabla_\lambda L(x^*, \lambda^*, \eta^*) - \alpha^* \\ &= -g(x^*) - \alpha^* \\ &= 0. \end{aligned}$$

Therefore

$$g(x^*) = -\alpha^* \leq 0$$

which is the first KKTb condition for the original program. Also,

$$(\alpha^*)^T \lambda^* = -(g(x^*))^T \lambda^* = -(\lambda^*)^T g(x^*) = 0$$

which is the third KKTb condition. The second KKTb condition is automatically fulfilled because

$$-\lambda^* \leq 0 \Rightarrow \lambda^* \geq 0.$$

All that remains is to show that KKTc holds. From the KKTa condition of (P),

$$\nabla_\eta L^P(\lambda^*, \eta^*, \alpha^*) = -\nabla_\eta L(x^*, \lambda^*, \eta^*) = -h(x^*) = 0$$

and so this condition holds true too.

We have seen that if the Saddle Inequality holds at a point (x^*, λ^*, η^*), then all the KKT conditions hold for that triple. Thus it is a KKT point and the theorem is proved.

We return to our prevous example:

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.t.} & -x_1 - x_2 + 4 \leq 0 \\ & x_1, x_2 \geq 0. \end{array}$$

The Lagrangian is

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda_1(-x_1 - x_2 + 4) + \lambda_2(-x_1) + \lambda_3(-x_2).$$

We have already seen that there is a KKT point at $x^* = (2, 2), \lambda^* = (4, 0, 0)$, and furthermore, $(2, 2)$ is a minimiser for the unconstrained function $L(x, (4, 0, 0))$.

Now consider the Lagrangian at x^*, considered as a function of λ only:

$$L(x^*, \lambda) = 4 - 2\lambda_2 - 2\lambda_3.$$

Maximising this over all $\lambda_1, \lambda_2, \lambda_3 \geq 0$ is easy: we obviously need $\lambda_2 = \lambda_3 = 0$.

This does not give us a value for λ_1, so we cannot solve for the KKT multipliers from this problem alone. However we note that for any value of λ_1, $L(x^*, \lambda)$ does not change, so we know that

$$L(x^*, \lambda) \leq L(x^*, (4, 0, 0)),$$

i.e. that the Saddle Inequality holds.

In fact, from the Theorem, noting that both sides of the Saddle Inequality holds actually proves that $((2,2),(4,0,0))$ is a KKT point!

Example.

$$\begin{aligned} \min \quad & x_1^2 + x_2 \\ \text{s.t.} \quad & x_2 \geq 0. \end{aligned}$$

The Lagrangian is

$$L(x_1, x_2, \lambda) = x_1^2 + x_2 + \lambda(-x_2) = x_1^2 + (1 - \lambda)x_2.$$

We look for all possible values which can solve the Saddle Inequality.

If λ^* has any other value besides 1, minimising the Lagrangian with respect to x_1 and x_2 will be an unbounded problem. So for the Saddle Inequality to hold, we need $\lambda^* = 1$.

Now we know that $L(x_1, x_2, \lambda^*) = x_1^2$. This clearly has a minimum at $x_1^* = 0$.

We also know that $\lambda^* = 1$ is the minimum of $L(x^*, \lambda)$, which is a linear function of λ, subject to $\lambda \geq 0$. The only way this can occur is if the coefficient of λ is itself 0 (note that this isn't the case if $\lambda^* = 0$). Therefore $-x_2 = 0$, which means $x_2^* = 0$.

Therefore the only point which satisfies the Saddle Inequality is $((0,0),1)$. From the Theorem, this is a KKT point, and also is the only KKT point.

Considering that solving the Saddle Inequality for even such a simple nonlinear program requires a couple of deductive reasonings (which would be hard to do in a general case), it is obvious that solving the Saddle Inequality directly is not a very good way to find the KKT points.

However, it does lead on to some interesting results, which make things much easier!