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 12, 2008

The Slater constraint qualification

We make a slight detour here to note another constraint qualification that can be applied to the KKT conditions to be sure that they hold. Recall that one of the constraint qualifications that we already know about requires $g(x)$ and $h(x)$ to be affine.

For a convex program, this can be relaxed to the Slater constraint qualification: this requires the program to be a convex program (in particular, $g(x)$ is convex and $h(x)$ is affine), and there must exist a strictly feasible point, i.e. a point x' such that $h(x') = 0$ and $g(x') < 0$.

Let's take a closer look at the Slater constraint qualification. Since $h(x)$ is affine, we can express it as

$$h(x) = Ax + a$$

where A is a $q \times n$ matrix and a is a vector of length q (q being the number of equality constraints).

In that case, for any point x, set $d = x' - x$. We know that

$$\nabla h(x)^T d = A(x' - x) = -a + a = 0$$

since both x and x' are feasible.

Now consider any constraint i that is active for x, i.e. $i \in I(x)$. We know that x lies on the line $g_i(x) = 0$. Because $g_i(x)$ is convex, the vector joining x to x' must lie entirely in the feasible region. This means that it is a descent direction for $g_i(x)$, i.e.

$$\nabla g_i(x)^T d = \nabla g_i(x)^T (x' - x) < 0.$$

These two conditions are just stating that the Mangasarian-Fromovitz qualification holds at every point. While this is not necessary for the Mangasarian-Fromovitz to hold (we just need it to hold at the supposed KKT point), it is certainly sufficient. Also the Slater qualification is easier to check!

This is just another demonstration of how having a convex program makes life easier for us.

Lagrangian duality

In the proof of the Lagrangian saddlepoint theorem, we noted that maximising the Lagrangian with respect to λ and η at the optimal x^* was equivalent to the following non-linear program:

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

We note that the objective function of this problem can be re-written as

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

Now we want to maximise this quantity over all $\lambda \geq 0$ and all η. But the only time that λ appears in this quantity is in the second term, and the only time that η appears in is the third term. Therefore we can write the problem as

$$\max_{\lambda \geq 0, \eta} L(x^*, \lambda, \eta) = f(x^*) + \max_{\lambda \geq 0} g(x^*)^T \lambda + \max_{\eta} h(x^*)^T \eta.$$

If we can solve this, then we will know the optimal KKT multipliers. The problem, as pointed out before, is that we do not know x^*, and thus cannot solve it directly.

Let us look at the problem when we do not know x^*, so we replace x^* by any given x.

The third term is $\max_\eta h(x)^T \eta$, which asks us to maximise a linear function in η. As we saw in previous examples, this is an unbounded problem unless the coefficient of η is 0, i.e. $h(x) = 0$.

The second term is $\max_{\lambda \geq 0} g(x)^T \lambda$, which again is a linear function. However, this time we have the constraint $\lambda \geq 0$. This will be an unbounded problem if any of the λ_i's have a positive coefficient.

Since the coefficient of λ_i is $g_i(x)$, the only way this problem will be unbounded is if $g_i(x) \leq 0$ for all i, i.e. if $g(x) \leq 0$. In this case, the optimum is at $\lambda = 0$ and so $g(x)^T \lambda = 0$.

We see from the above that the program is unbounded if and only if x is an infeasible point. Therefore we write the program as a function of x:

$$\phi(x) = \max_{\lambda \geq 0, \eta} L(x, \lambda, \eta) = \begin{cases} f(x) & \text{if } x \text{ is feasible for (NLP)} \\ \infty & \text{otherwise.} \end{cases}$$

This is actually a penalty function!

In fact, looking closer reveals that it is the *ideal* penalty function: when x is feasible, it takes the value of $f(x)$, and when x is infeasible, it is ∞, which excludes any minimiser from choosing an infeasible point.

Therefore, we can solve the original non-linear program by minimising this function, in an unconstrained manner:

$$\min_x \phi(x) = \min_x \max_{\lambda \geq 0, \eta} L(x, \lambda, \eta).$$

In other words, we can maximise the Lagrangian with respect to λ and η (ensuring the left-hand side of the Saddle Inequality), while keeping x arbitrary, and then minimise the maximum with respect to x (ensuring the right-hand side of the Saddle Inequality) to obtain the solution.

This gives us another idea: we can do this in reverse! We can minimise the Lagrangian with respect to x, keeping the multipliers arbitrary, and then maximise the minimum with respect to λ and η to obtain the solution.

In other words, we exchange the minimum and the maximum:

$$\min_x \max_{\lambda \geq 0, \eta} L(x, \lambda, \eta) = \max_{\lambda \geq 0, \eta} \min_x L(x, \lambda, \eta) = \max_{\lambda \geq 0, \eta} \psi(\lambda, \eta)$$

where the last equality is the definition of $\psi(\lambda, \eta)$.

This gives us two new problems which should have the same solution as the original problem — the penalty formulation of the original NLP,

$$\min_{x} \phi(x),$$

and the *Lagrangian dual* problem,

$$\max_{\lambda \geq 0, \eta} \psi(\lambda, \eta).$$

Example. Let us return to one of our previous examples:

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

First we look at the penalty formulation of the primal problem:

$$
\begin{aligned}
\phi(x) &= \max_{\lambda \geq 0, \eta} L(x, \lambda, \eta) \\
&= \max_{\lambda \geq 0} x_1^2 + x_2^2 + \lambda_1(-x_1 - x_2 + 4) + \lambda_2(-x_1) + \lambda_3(-x_2) \\
&= x_1^2 + x_2^2 + \max_{\lambda_1 \geq 0}(-x_1 - x_2 + 4)\lambda_1 + \max_{\lambda_2 \geq 0} -x_1\lambda_2 + \max_{\lambda_3 \geq 0} -x_2\lambda_3.
\end{aligned}
$$

If any of the constraints are violated, this is an unbounded problem, so $\phi(x) = \infty$ for any infeasible point. Apart from that, however, it is not immediately obvious how to solve the maximisation problem to express $\phi(x)$ in terms of x only (so that we can apply a unconstrained minimiser).

Now consider the Lagrangian dual:

$$\begin{aligned}
\psi(\lambda) &= \min_x L(x,\lambda) \\
&= \min_x x_1^2 + x_2^2 + \lambda_1(-x_1 - x_2 + 4) + \lambda_2(-x_1) + \lambda_3(-x_2) \\
&= \min_x 4\lambda_1 + (x_1^2 - (\lambda_1 + \lambda_2)x_1) + (x_2^2 - (\lambda_1 + \lambda_3)x_2) \\
&= 4\lambda_1 + \min_{x_1}(x_1^2 - (\lambda_1 + \lambda_2)x_1) + \min_{x_2}(x_2^2 - (\lambda_1 + \lambda_3)x_2).
\end{aligned}$$

Because this is not a linear function in x, we can actually solve it.

We note that for any a, the function $y^2 - ay$ is a quadratic function, and its minimum occurs at the turning point $y = \frac{a}{2}$. Therefore its minimum value is $\left(\frac{a}{2}\right)^2 - a\frac{a}{2} = -\frac{1}{4}a^2$. We substitute this in the above to derive

$$\psi(\lambda) = 4\lambda_1 - \frac{1}{4}(\lambda_1 + \lambda_2)^2 - \frac{1}{4}(\lambda_1 + \lambda_3)^2.$$

Therefore the Lagrangian dual problem is

$$\begin{array}{ll} \max & 4\lambda_1 - \frac{1}{4}(\lambda_1 + \lambda_2)^2 - \frac{1}{4}(\lambda_1 + \lambda_3)^2 \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

This is not an unconstrained problem (so we cannnot apply unconstrained methods), but it does have very simple constraints, making it easy to solve. The solution can be seen to be $\lambda^* = (4, 0, 0)$, our optimal KKT multipliers.

Furthermore, at the point $\lambda^* = (4, 0, 0)$, the Lagrangian dual problem has the optimal function value

$$4 \times 4 - \frac{1}{4}(4 + 0)^2 - \frac{1}{4}(4 + 0)^2 = 8,$$

and at the point $x^* = (2, 2) = \min_x L(x, (4, 0, 0))$, the original program has the optimal function value

$$2^2 + 2^2 = 8.$$

This shows that the two problems have identical optimal function values.

However, the original program is a minimising problem, and the Lagrangian dual is a maximising problem. In particular, this means that for any feasible point x, $f(x) \geq 8$, and for any feasible λ for the Lagrangian dual, $\psi(\lambda) \leq 8$. Therefore, for any feasible pair of points x, λ, we know that

$$\psi(\lambda) \leq f(x).$$

The example above demonstrates the following duality theorem.

Theorem. Let (NLP) be a convex program, and let $\phi(x)$ and $\psi(\lambda, \eta)$ be defined as previously.

1. (Weak Lagrangian duality) Let x be feasible for (NLP) and the multipliers λ, η be such that $\lambda \geq 0$. Then $\psi(\lambda, \eta) \leq \phi(x)$.
2. (Strong Lagrangian duality) A triple (x^*, λ^*, η^*) is a KKT point of (NLP) if and only if $\lambda^* \geq 0$ and $\psi(\lambda^*, \eta^*) = \phi(x^*)$.