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 14, 2008

Proof.

1. From the definitions,

$$\psi(\lambda, \eta) = \min_{x'} L(x', \lambda, \eta) \leq L(x, \lambda, \eta) \leq \max_{\lambda' \geq 0, \eta'} L(x, \lambda', \eta') = \phi(x).$$

2. Suppose (x^*, λ^*, η^*) is a KKT point. Then the Saddle Inequality holds. Therefore

$$\psi(\lambda^*, \eta^*) = \min_{x} L(x, \lambda^*, \eta^*) = L(x^*, \lambda^*, \eta^*)$$

   from the Saddle Inequality.

Furthermore

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

again from the Saddle Inequality. This means that $\psi(\lambda^*, \eta^*) = \phi(x^*)$.

Now suppose that $\psi(\lambda^*, \eta^*) = \phi(x^*)$. From the proof of part 1,

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

so all three quantities are equal.

Therefore

$$\min_{x} L(x, \lambda^*, \eta^*) = L(x^*, \lambda^*, \eta^*)$$

which implies that x^* minimises $L(x, \lambda^*, \eta^*)$. Similarly

$$\max_{\lambda \geq 0, \eta} L(x^*, \lambda, \eta) = L(x^*, \lambda^*, \eta^*)$$

which implies that λ^* and η^* maximise $L(x^*, \lambda, \eta)$ (subject to $\lambda \geq 0$).

However, this is the Saddle Inequality, so (x^*, λ^*, η^*) is a KKT point and the theorem is proved.

Example. Consider the non-linear program

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

The Lagrangian is

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

The penalty function is

$$\phi(x) = \max_{\lambda \geq 0} x_1^2 + x_2 + \lambda(2 - x_2).$$

The objective function for the Lagrangian dual is

$$\psi(\lambda) = \min_{x} x_1^2 + x_2 + \lambda(2 - x_2).$$

Now for any feasible x and $\lambda \geq 0$, weak duality tells us that

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

Now consider the point $(x, \lambda) = ((0,2), 1)$. The two functions are

$$\phi(0,2) = \max_{\lambda \geq 0} 0^2 + 2 + 0\lambda = 2,$$

and

$$\psi(\lambda) = \min_{x} x_1^2 + x_2 + 2 - x_2 = \min_{x_1} x_1^2 + 2 = 2.$$

By strong duality, this tells us that $((0,2),1)$ is a KKT point of the program.

Note that from weak Lagrangian duality,

$$\max_{\lambda \geq 0, \eta} \psi(\lambda, \eta) \leq \min_{x} \phi(x),$$

so if the conditions for strong duality hold, it is easy to see that (x^*, λ^*, η^*) is not just a KKT point, but also a global minimum. This also follows from the fact that the program is convex.

We can also see from the above inequality that in this case, (λ^*, η^*) is a global maximum for the Lagrangian dual.

Weak and strong duality also offer some other interesting results.

Corollary. If we set z to be the optimal function value of the non-linear program,

$$z = \min_x \{f(x) : g(x) \leq 0, (x) = 0\},$$

then if the program is a convex program and $\lambda \geq 0$, we have

$$z \geq \psi(\lambda, \eta)$$

for any η.

This tells us that the optimal function value of the original program is bounded below by the objective of the Lagrangian dual, evaluated at any feasible point.

This is useful, because we can then evaluate the Lagrangian dual at any feasible point to gain a lower bound for the original program, even if we do not know what the solution either program is. By evaluating the dual at different feasible points, we can (hopefully) get closer and closer to the actual optimum value for the primal program.

Proof. From weak duality, we know that at any feasible (for the Lagrangian dual) λ and η,

$$\min_x \phi(x) \geq \psi(\lambda, \eta).$$

But $\min_x \phi(x)$ is the optimal objective function of the original program (since $\phi(x)$) is the ideal penalty function.

Example. Going back to our first example:

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

The Lagrangian dual objective function is

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

Evaluating this function at any feasible λ gives a lower bound for the objective function of the primal. So if the optimal value of the primal is z, we can say, for example, that

$$\begin{aligned} z &\geq \psi(0,0,0) = 0 \\ z &\geq \psi(1,1,1) = 2 \\ z &\geq \psi(1,2,3) = -\frac{9}{4} \\ z &\geq \psi(4,0,0) = 8. \end{aligned}$$

We already know that $z = 8$, and so can see that these inequalities are valid.

Corollary. If x is feasible for a convex program, $\lambda \geq 0$, and $f(x) = \psi(\lambda, \eta)$, then (x, λ, η) is a KKT point and x is a global minimum of the program.

This works in much the same way as strong duality, except that we need merely to evaluate the original objective function (and the Lagrangian dual).

Proof. If x is feasible, then

$$\begin{aligned}\phi(x) &= \max_{\lambda\geq 0,\eta} L(x,\lambda,\eta) \\ &= \max_{\lambda\geq 0,\eta} f(x) + g(x)^T\lambda + h(x)^T\eta \\ &= f(x) + \max_{\lambda\geq 0} g(x)^T\lambda \\ &= f(x).\end{aligned}$$

The last equation follows because $g(x) \leq 0$. Therefore (x,λ,η) satisfies the conditions for strong duality and is a KKT point. Since the program is convex, it is also a global minimum.

Example. In the previous example, we know that $\psi(4, 0, 0) = 8$. Evaluating the objective function at the point $(2, 2)$ gives

$$f(2, 2) = 2^2 + 2^2 = 8$$

which demonstrates that $(2, 2)$ is the global minimum of the original program.

Wolfe duality

The Lagrangian dual has some good properties but it does have a drawback. Recall that the objective function is

$$\psi(\lambda, \eta) = \min_x L(x, \lambda, \eta).$$

This involves minimising the Lagrangian just to calculate the objective function. If the Lagrangian is complex, this can be difficult and undesirable.

The Wolfe dual gets around this difficulty by observing that since the above problem is an unconstrained problem, at the minimum of the Lagrangian its gradient must be 0 with respect to the decision variables:

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

At this point, $\psi(x, \lambda, \eta)$ is equal to the Lagrangian, which allows us to convert the Lagrangian dual into the Wolfe dual.

The Wolfe dual is

$$\begin{array}{ll} \max_{x,\lambda,\eta} & L(x,\lambda,\eta) \\ \text{s.t.} & \lambda \geq 0 \\ & \nabla_x L(x,\lambda,\eta) = 0. \end{array}$$

Note that this is a (constrained) optimisation problem in all three variable sets, x, λ, and η. As with the Lagrangian dual, it is a maximisation problem.

We make two further observations. Firstly, we cannot assume that $\nabla_x L(x, \lambda, \eta)$ is affine, and therefore the Wolfe dual may not be a convex program.

Also, note that a feasible point in the primal has to satisfy $g(x) \leq 0$, $h(x) = 0$, but these constraints are not applied to x in the Wolfe dual. The equality constraint on x in the dual is not in the primal either, so there is no relationship between primal and dual feasible points.

Let us look at the Wolfe dual of a linear program:

$$\begin{array}{ll} \text{min} & c^T x \\ \text{s.t.} & x \geq 0 \\ & Ax = b. \end{array}$$

If we take $g(x) = -x$ and $h(x) = b - Ax$, the Lagrangian is

$$\begin{aligned} L(x, \lambda, \eta) &= c^T x - \lambda^T x + \eta^T (b - Ax) \\ &= (c - \lambda - A^T \eta)^T x + \eta^T b. \end{aligned}$$

This gives us $\nabla_x L(x, \lambda, \eta) = c - \lambda - A^T \eta$, so the Wolfe dual of this program is

$$\begin{array}{ll} \max_{x,\lambda,\eta} & (c - \lambda - A^T\eta)^T x + \eta^T b \\ \text{s.t.} & \lambda \geq 0 \\ & c - \lambda - A^T \eta = 0. \end{array}$$

Now, the equality constraint shows that the coefficient of x in the objective function is 0. This removes all of the x's from the dual. Furthermore, if we write the equality constraint as $\lambda = c - A^T \eta$, we can then substitute this into the inequality constraint to remove λ.

This results in a simplified Wolfe dual of

$$\begin{array}{ll} \max_{\eta} & b^T \eta \\ \text{s.t.} & A^T \eta \leq c. \end{array}$$

This is the standard dual of the original linear program.

Example. We return to our regular example:

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

The Lagrangian of this program is

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

Therefore the Wolfe dual is

$$\begin{array}{ll} \max & x_1^2 + (-\lambda_1 - \lambda_2)x_1 + x_2^2 + (-\lambda_1 - \lambda_3)x_2 + 4\lambda_1 \\ \text{s.t.} & \lambda \geq 0 \\ & 2x_1 - \lambda_1 - \lambda_2 = 0 \\ & 2x_2 - \lambda_1 - \lambda_3 = 0. \end{array}$$

Again, the Wolfe dual can be simplifed somewhat because of the equality constraints.

From the equality constraints, we know that $-\lambda_1 - \lambda_2 = -2x_1$ and $-\lambda_1 - \lambda_3 = -2x_2$. This eliminates λ_2 and λ_3 from the dual, which gives:

$$\begin{array}{ll} \max & -x_1^2 - x_2^2 + 4\lambda_1 \\ \text{s.t.} & \lambda_1 \geq 0 \\ & 2x_1 - \lambda_1 \geq 0 \\ & 2x_2 - \lambda_1 \geq 0. \end{array}$$

This can be solved using the KKT conditions to generate the optimal solution $x_1 = 2, x_2 = 2, \lambda_1 = 4$. These are the optimal values for the primal as well, and back substitution gives $\lambda_2 = 2x_1 - \lambda_1 = 0$ and $\lambda_3 = 2x_2 - \lambda_1 = 0$.