Lagrange Duality Example

16:198:536

A frequently useful trick in optimization theory is that of *duality* - this essentially amounts to a change of variables to re-express the original problem in frequently useful ways. As an example of this, consider the 'primal' problem

$$\min_{\underline{x}} f(\underline{x}) \quad \text{(s.t.)} \quad \forall i = 1, \ldots, m : 0 \geq g_i(\underline{x}), \tag{1}$$

where $\underline{x}$ is a d-dimensional real valued vector, subject to m-many inequality constraints of the form $0 \geq g_i(\underline{x})$. That is, we want to solve for d-many unknowns to optimize the function f subject to the m-many constraints.

The principle of **Lagrangian Duality** says that under some conditions on f and g_i, the solution to the above is equivalent to the solution to

$$\max_{\underline{\lambda}} \left[\min_{\underline{x}} f(\underline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\underline{x}) \right] \quad \text{(s.t.)} \quad \forall i : \lambda_i \geq 0, \tag{2}$$

where we have no constraints on $\underline{x}$, and have introduced variables (Lagrange multipliers), one for each inequality constraint.

Having eliminated the constraints in this way, the interior minimum occurs when the gradient with respect to $\underline{x}$ (the vector of all the derivatives) is 0, i.e.,

$$\nabla f(\underline{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\underline{x}) = 0. \tag{3}$$

If we can solve for the above system (which will frequently require expressing $\underline{x}$ in terms of the λ_i, we can essentially flip the problem on its head and think about maximizing with respect to the λ_i. One of the values of this for instance is that the number of variables in the new problem will be m, rather than d. So if the number of constraints is much smaller than the total number of variables in the the 'primal' problem, then this can significantly decrease the complexity of the problem.

As an example of this, consider trying to solve

$$\min_{\underline{x}} \frac{1}{2} ||\underline{x} - \underline{a}||^2 \quad \text{(s.t.)} \quad ||\underline{x}||^2 \leq 1. \tag{4}$$

Essentially, the above is asking us to find the d-dimensional vector $\underline{x}$ that is closest to some given point $\underline{a}$, but is itself within the unit sphere. Geometrically, if $\underline{a}$ is in the unit sphere, the answer will be $\underline{x}^* = \underline{a}$, but if $\underline{a}$ is outside the unit sphere, then the answer will be given by $\underline{x}^*$ as $\underline{a}/||\underline{a}||$, the unit vector pointing in the direction of $\underline{a}$ - the closest point on the unit sphere to $\underline{a}$.

Expressing this in terms of Lagrangian duality, we want

$$\max_{\lambda > 0} \left[\min_{\underline{x}} \frac{1}{2} ||\underline{x} - \underline{a}||^2 + \lambda(||\underline{x}||^2 - 1) \right]. \tag{5}$$

Solving the interior minimum, we want

$$\nabla \left[\frac{1}{2} ||\underline{x} - \underline{a}||^2 \right] + \lambda \nabla \left[||\underline{x}||^2 - 1 \right] = 0, \tag{6}$$

or

$$(\underline{x} - \underline{a}) + 2\lambda \underline{x} = 0, \tag{7}$$

or

$$\underline{x} = \frac{1}{1 + 2\lambda} \underline{a}. \tag{8}$$

This re-expresses the solution $\underline{x}^*$ in terms of the new unknown variable λ. Notice that we have exchanged a d-dimensional problem for a 1-dimensional problem, finding λ. Substituting this into the original problem, we now have

$$\max_{\lambda>0}\left[\frac{1}{2}||\frac{1}{1+2\lambda}\underline{a}-\underline{a}||^2+\lambda(||\frac{1}{1+2\lambda}\underline{a}||^2-1)\right]. \tag{9}$$

This simplifies somewhat:

$$\max_{\lambda>0}\left[\frac{1}{2}\left(\frac{1}{1+2\lambda}-1\right)^2||\underline{a}||^2+\frac{\lambda}{(1+2\lambda)^2}||\underline{a}||^2-\lambda\right], \tag{10}$$

or

$$\max_{\lambda>0}\left[\frac{\lambda}{1+2\lambda}||\underline{a}||^2-\lambda\right]. \tag{11}$$

Which, you'll observe, is a relatively simple optimization problem in one variable, albeit a variable constrained to be positive.

To complete the solution, it's worth noting that the derivative of the objective function is given by $||\underline{a}||^2/(1+2\lambda)^2-1$, which equals 0 when $\lambda=(||\underline{a}||-1)/2$ - which is only non-negative if $||\underline{a}||\geq 1$! Otherwise the maximum occurs when $\lambda=0$.

Substituting these in to the original problem: we get that $\underline{x}^*=\underline{a}/(1+2\lambda^*)=\underline{a}/||\underline{a}||$, when $\underline{a}$ is outside the unit sphere, or $\underline{x}^*=\underline{a}/(1+2\lambda^*)=\underline{a}$ when $\underline{a}$ is within the unit sphere, exactly the answer anticipated.