**Lagrange duality: an (usually) important step in optimization**

Consider an optimization problem in standard form:

(0-1)

We do not assume the problem (0-1) is convex!

Basic idea of Lagrange duality: take the constraints in (0-1) into account in the objective function. Define the Lagrangian *L*:

(0-2)

Note Lagrangian by itself has no constraints. The elements in the vectors λ and ν are called Lagrange multipliers.

We define a Lagrangian dual function:

(0-3)

For a fixed *x*, is an affine (linear) combination of λ and ν, so it is concave. Note in formal expressions people use infimum instead of minimum but here we just use minimum for easier understanding.

If we ensure that , then for a feasible point (feasible means the point satisfies the constraints in (0-1))we can be sure that . In addition, we can also easily show that

. (0-4)

Keep in mind of this inequality, and the constraints . The inequality has an important implication, that for an arbitrary , sets a lower bound of the original object. Naturally, we can introduce the Lagrangian dual problem:

(0-5)

The above problem is convex, because of convexity in and the linearity in constraints.

Keep in mind, that the dual problem can be very useful in many ways. For example, if Slater’s condition is met, the primal problem and the dual problem has no gap. As a result, if in (0-5), is easily solved analytically, the primal problem can be easily solved. The dual problem also has been widely used for sensitivity analysis.

Here we are only concerned about the certificate of suboptimal and stopping criteria that dual problem brings about. In (0-5), each feasible pair (λ, ν) brings sets a lower bound on the optimal value. Mathematically speaking, for each feasible set of (x, λ, ν):

(0-6)

As a result, even if we do not know the actual optimality , we can still establish an “overestimate” of convergence such that if

(0-7)

We would know the accuracy is good enough.

(However, keep in mind that if strong duality cannot be guaranteed, the convergence measure cannot be infinitely small, due to existence of duality gap!)

**KKT conditions**

Assumption: zero duality gap, but no assumption about convexity yet!

If there is a set of optimal points for both problems , we must (necessary) have:

Note the last equation might be weird at first sight. However it is necessary because if for any non-zero , x could always be arbitrarily adjusted in (0-5) and result in .

Also note for a convex primal problem, the KKT conditions are not only necessary but also sufficient for optimal!

**Convex optimization algorithms hierarchy**

1. Linear equality constrained quadratic problems – direct, analytical solution
2. Linear equality constrained optimization problems (twice differentiable objective) – Newton’s method, reduced into a series of 1.
3. Linear equality and inequality constrained optimization problems – interior point method, reduced into a series of 2.

**General form of constrained optimization**

(1)

Assumptions:

1. is convex, are linear, A is non-singular.
2. Problem is solvable (optimal exists).
3. Problem is strictly feasible ( that satisfies the constraints).

**Logarithmic barrier function and central path**

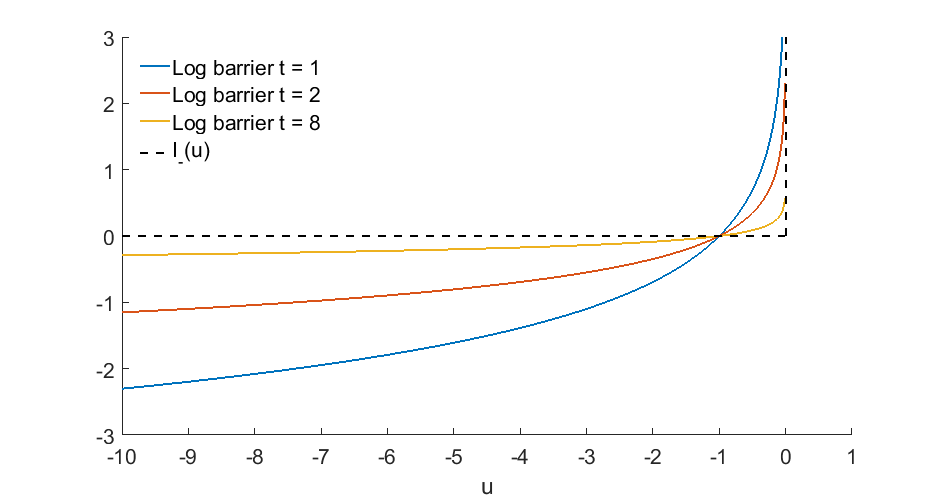
We can rewrite the forms of (1) into:

(2)

The basic idea behind the barrier method is replacing the function by a log barrier function

(3)

We call the above function “log barrier”. It is convex and non-decreasing. The advantage of it over the original is that that log barrier function is differentiable!



As upper plot shows, intuitively we truly see that as t gets larger, the log barrier function gets more similar to .

The modified optimization function is then

(4)

Define the following equation for easier discussion (call the log function hereafter):

(5)

Also, the gradient and Hessian of the log barrier function are:

(6)

(7)

Now, consider the details in (4):

If a solution *x\*(t)* can be found for a value *t*, we would be sure that the point *x\*(t)* satisfies the constraints, and there exists a vector *v\** such that

(8)

A simple reasoning for the above equation, is that for an optimization problem with equality constraints, the object is optimized only when the gradients of the object and the constraints are parallel!!!

The equation (8) must be carefully compared with the KKT conditions directly in order for people to fully appreciate the design of the log barriers. Basically, (8) forms the SAME form as that in KKT condition, if we let

(9)

Because the point *x\*(t)* satisfies the constraints, the polarity of meets the second KKT condition, and can be derived

(10)

Here *m* is the dimension of constraints. The above equation, together with (0-6), imply that we naturally get a measure of convergence which is m/t. In other words, mathematically we proved the convergence of the problem, as t increases!

**Barrier method**

The barrier method can be easily established.

Given *x*, *t* = *t0* > 0, *µ*>1, *ϵ*>0:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Repeat (outer loop)

1. “Centering step”: Starting from *x*, compute *x\*(t)* by minimizing *tf0(x)* + , s.t. *Ax = b*. This is solved iteratively by Newton’s method (inner loop).
2. Update: *x* := *x\*(t)*
3. Stopping criterion: quit if *m/t* < *ϵ*
4. Increase *t*: *t*:= *µt*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Note the number of *µ* has a tradeoff: if too small, many outer iterations happen. If too big, many inner iterations happen.

Note: people might also think why not directly start from a large *t*? Because using a large *t*, the log barrier function would change very quickly near optimal. It is in fact possible when the dimension of problem is small. However, when the dimension is large interior points is better off.