Optimization Notes

Luis Zermeño

September 4, 2012

1 Unconstrained optimization

Suppose we want to solve

$$\max_{x \in \mathbb{R}^n} \quad f(x;a)$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable (vector x is a vector of variables and a is a vector of parameters). We know that necessary conditions for $x^*(a)$ to be a maximum are:

$$\frac{\partial f(x^*(a);a)}{\partial x_i} = 0 \;\; \forall \;\; i = 1, ..., n$$

These conditions are not sufficient since we might be finding a local minimum, a local maximum or an inflection point. Any $x^*(a)$ that satisfies these conditions is called a **critical point**.

1.1 Envelope Theorem

For all $a \in \mathbb{R}^m$, let $x^*(a)$ be a critical point of $f(x;a)$. We are interested in analyzing how our objective function evaluated at these critical points varies as a function of a. Let $V(a) \equiv f(x^*(a);a)$. A simple version of the envelope theorem states that, if $x^*(a)$ is differentiable, then

$$\frac{\partial V(a)}{\partial a_j} \equiv \underbrace{\frac{\partial f(x^*(a);a)}{\partial a_j}}_{\text{Direct Effect}} + \underbrace{\sum_{i=1}^{n} \frac{\partial f(x^*(a);a)}{\partial x_i}\frac{\partial x_i^*(a)}{\partial a_j}}_{\text{Indirect Effect}\equiv 0} \equiv \underbrace{\frac{\partial f(x^*(a);a)}{\partial a_j}}_{\text{Direct Effect}},$$

where the equality follows from the conditions that characterize a critical point.

2 Constrained optimization

Now the problem is

$$\max_{x\in\mathbb{R}^n} \quad f(x;a)$$

Subject to

$$g_j(x;a) = b_j, \quad j = 1, ..., J$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $j = 1, ..., J$, are continuously differentiable functions, and $b_j \in \mathbb{R}$ for all j. We assume that the constraint set is non-empty.

Define the Lagrangian function, $L : \mathbb{R}^n \times \mathbb{R}^J \times \mathbb{R}^m \times \mathbb{R}^J \to \mathbb{R}$, as:

$$L(x, \lambda; a, b) \equiv f(x;a) - \sum_{j=1}^{J} \lambda_j (g_j(x;a) - b_j),$$

where a and b are vectors of parameters.

Key result 1: Under fairly general conditions, if $x^*(a, b)$ is a solution of the constrained problem (given a), then there exists $\lambda^*(a, b)$ such that $(x^*(a, b), \lambda^*(a, b))$ is a critical point of $L(x, \lambda; a, b)$. That is,

$$\frac{f(x^*(a,b);a)}{\partial x_i} = \sum_{j=1}^{J} \lambda_j^*(a) \frac{\partial g_j(x^*(a,b);a)}{\partial x_i} \quad \forall \ \ i = 1, .., n; \tag{1}$$

$$g_j(x^*(a,b);a) = b_j \quad \forall \ \ j = 1, ..., J. \tag{2}$$

In other words, if $x^*(a,b)$ is a solution to our problem, it must satisfy (2) (which means that it must be feasible given the constraints) and (1), which at first sight is a little harder to understand. Note that (1) can be rewritten as

$$\nabla f(x^*(a,b);a) = \sum_{j=1}^{J} \lambda_j \nabla g_j(x^*(a,b);a),$$

which says that the gradient of $f(x;a)$ must be a linear combination of the gradients of the constraints at $x^*(a,b)$. This condition is particularly intuitive when we have one constraint and $n=2$, because then all it says is that the slope of the constraint at $x^*(a,b)$ must coincide with the slope of the level curve of $f(x;a)$ at that point.

When does the result fail? When the slopes are not well defined at the optimum. For example, suppose that the problem is,

$$\max_{x,y} \quad -y$$

s.t.

$$y^3 = x^2$$

It is clear that the solution is $x=y=0$. However, the Lagrangian in this case has no critical points. Indeed, we would have to find x, y, λ such that

$$-1 = 3y^2\lambda$$

$$2x\lambda = 0$$

$$y^3 = x^2$$

which can never hold simultaneously. The problem is that, if we plot the constraint, it has a kink at (0,0), so its slope is not well defined.

2.1 When are the Lagrange conditions sufficient?

The previous result states that, under quite general conditions, if $x^*(a, b)$ solves our problem, then it must be part of a critical point of the Lagrangian. Is the converse true? Not in general. For example, suppose our problem is

$$\max_{x,y} \quad x^2 + y^2,$$

s.t.

$$x + y = 1.$$

It is easy to verify that $x = \frac{1}{2}$, $y = \frac{1}{2}$ and $\lambda = 1$ is a critical point of the Lagrangian. However, it this problem does not have a solution (we can make x arbitrarily large and y arbitrarily small and get the value to infinity). The problem is that the objective is not quasi-concave.

Key result 2: Given a, b, let $f(x; a)$ be **quasi-concave**, and the constraint set, $\{x \in \mathbb{R}^n \mid g_j(x; a) = b_j \;\; \forall \;\; j\}$, be **convex**. Suppose that there is a pair (x^*, λ^*) that satisfies the Lagrange conditions, and that at least one of the following conditions is satisfied:

1. $\nabla f(x^*; a) \neq 0$ and $f(x; a)$ is twice differentiable;
2. $f(x; a)$ is a concave function.

Then x^* solves our problem given a, b.

2.2 Interpretation of λ (the Lagrange multipliers)

For all (a, b), suppose that $(x^*(a, b), \lambda^*(a, b))$ is a critical point of the Lagrangian, and that $x^*(a, b)$ solves our constrained problem. Note that the value function of our problem, $V(a, b)$, satisfies:

$$V(a, b) \equiv L(x^*(a, b), \lambda^*(a, b); a, b) \equiv f(x^*(a, b)) - \sum_{j=1}^{J} \lambda_j^*(b)(g_j(x^*(a, b)) - b_j).$$

Thus, using the envelope theorem,

$$\frac{\partial V(a,b)}{\partial b_j} \equiv \frac{\partial L(x^*(b), \lambda^*(b), b)}{\partial b_j} \equiv \lambda_j^*(a,b).$$

In words, the Lagrange multiplier, $\lambda_j(a,b)$, tells us how the value function of the problem changes after a marginal increase in b_j when we start from the point (a,b).

3 Optimization with inequality constraints and the Kuhn-Tucker method

Now we are dealing with the problem

$$\max_{x\in\mathbb{R}^n} \quad f(x;a)$$

Subject to

$$g_j(x;a) = b_j, \quad j = 1, ..., J$$

$$h_k(x;a) \geq c_k, \quad k = 1, ..., K$$

This problem is solved by turning the inequality constraints "on" and "off", and solving the problem only with the constraints that are "on", with the method that was described above. Thus, unfortunately, in these problems we typically have to consider different cases. The Kuhn-Tucker method is based on these observations. Now we define the Lagrangian as

$$L(x, \lambda, \mu; a, b, c) = f(x;a) - \sum_{j=1}^{J} \lambda_j (g_j(x;a) - b_j) + \sum_{k=1}^{K} \mu_k (h_k(x;a) - c_k).$$

We say that x^*, λ^*, μ^* satisfy the Kuhn-Tucker conditions (given a, b, c) if:

$$\frac{f(x^*;a)}{\partial x_i} = \sum_{j=1}^{J} \lambda_j^* \frac{\partial g_j(x^*;a)}{\partial x_i} - \sum_{k=1}^{K} \mu_k^* \frac{\partial h_l(x^*;a)}{\partial x_i} \quad \forall \;\; i = 1, .., n$$

$$g_j(x^*;a) = b_j \quad \forall \;\; j = 1, ..., J$$

$$\mu_k^* \geq 0, \quad h_k(x^*; a) - c_k \geq 0, \quad \mu_k^*(h_k(x^*; a) - c_k) = 0 \quad \forall \;\; k = 1, ..., K$$

The third set of conditions reflects the idea that we are only turning constraints "on" and "off". For example, if we think that the only constraint that is not binding is constraint k, then we want to ignore it and check that it is satisfied at the end. Thus, we make $\mu_k^* = 0$, and conjecture that all the other constraints must be satisfied with equality. That is, for $l \neq k$ we make $h_l(x^*; a) - c_l = 0$. Note that we end up with almost exactly the same conditions as those that characterized the critical points of the Lagrangian of the problem with only equality constraints. The only difference is that now, in order for the critical points to be candidates for the optimum, they have to satisfy in addition $h_k(x^*; a) - c_k \geq 0$ (constraint k was indeed not binding), and $\mu_l^* \geq 0$ for all $l \neq k$. The intuition of why we have to add the non-negativity constraints for the multipliers is simple if we understand their interpretation. If constraint l is binding, it has to be the case that the value function of the problem, $V(a, b, c)$, decreases when c_l increases (the constraint set is shrinking). Hence, from the envelope theorem,

$$\frac{\partial V(a, b, c)}{\partial c_l} \equiv \frac{\partial L(x^*(a, b, c), \lambda^*(a, b, c), \mu^*(a, b, c); a, b, c)}{\partial c_l} \equiv -\mu_l^*(a, b, c) \leq 0,$$

which holds if and only if $\mu_l^* \geq 0$. This constraint has to come from "outside" because it reflects an insight that comes from the direction of the inequality.

When are the Kuhn-Tucker conditions necessary and sufficient? Key Results 1 and 2 extend to this case, since the logic is exactly the same.

3.1 Kuhn-Tucker in practice

Example 1:

$$\max_{x,y} -(x - a)^2 - (y - b)^2$$

Subject to

$$x + y \leq 1$$

$$x, y \geq 0$$

The Lagrangian is

$$L(x, y, \lambda, \mu_x, \mu_y) = -(x - a)^2 - (y - b)^2 - \lambda(x + y - 1) + \mu_x x + \mu_y y$$

The Kuhn Tucker conditions are:

$$-2(x - a) = \lambda - \mu_x$$

$$-2(y - b) = \lambda - \mu_y$$

$$\lambda \geq 0, \quad x + y \leq 1, \quad \lambda(x + y - 1) = 0$$

$$\mu_x \geq 0, \quad x \geq 0, \quad \mu_x x = 0$$

$$\mu_y \geq 0, \quad y \geq 0, \quad \mu_y y = 0$$

1. Suppose that only the first constraint is active. This implies $\mu_x = \mu_y = 0$ and $x+y = 1$. Thus,

$$x - a = y - b$$

$$x + y = 1$$

$$y = \frac{1 + b - a}{2}$$
$$x = \frac{1 + a - b}{2}$$

We need to verify that $\lambda \geq 0$, and $x, y \geq 0$. All three will hold iff $b \geq a - 1$, $b \leq a + 1$, and $a \geq x \iff a + b \geq 1$.

2. Suppose only $x = 0$ binds. This implies that $\lambda = \mu_y = x = 0$. Then $y = b$. We need to verify that $x + y \leq 1 \iff b \leq 1$, $y \geq 0 \iff b \geq 0$, and $\mu_x \geq 0 \iff a \leq 0$.

3. The case where only $y = 0$ binds is analogous.

4. Suppose that non of the constraints binds. Then, $x = a$, $y = b$ and $\lambda = \mu_x = \mu_y = 0$. The constraints will indeed be all satisfied iff (a, b) is in the constraint set.

5. Constraints $x = 0$ and $y = 0$ hold. Then $\lambda = 0$. Since the third constraint holds trivially, we only need to verify that $\mu_x \geq 0 \iff a \leq 0$ and $\mu_y \geq 0 \iff b \leq 0$.

6. $x = 0$ and $x + y = 1$ bind. Then $y = 1$ so $\mu_y = 0$. We need to verify that $\mu_x \geq 0$ and $\lambda \geq 0$. We have

$$2b - 2 = \lambda$$

$$2a = \lambda - \mu_x$$

Thus,

$$\lambda \geq 0 \iff b \geq 1$$

$$\mu_x \geq 0 \iff (b - a) - 1 \geq 0 \iff b \geq a + 1$$

7. The case where $y = 0$ and $x + y = 1$ bind is analogous.

Example 2 (Exercise):

$$\max_{x,y} \quad 2y^{\frac{1}{2}} + x$$

s.t.

$$Ax + y = c$$

$$x, y \geq 0$$