Constrained Optimization

Joshua Wilde, revised by Isabel Tecu, Takeshi Suzuki and María José Boccardi

August 13, 2013

1 General Problem

Consider the following general constrained optimization problem:

$$\max_{x_i \in \mathbb{R}} f(x_1, \ldots, x_n) \text{ subject to :}$$

$$g_1(x_1, \ldots, x_n) \le b_1, \ldots, g_k(x_1, \ldots, x_n) \le b_k,$$
$$h_1(x_1, \ldots, x_n) = c_1, \ldots, h_m(x_1, \ldots, x_n) = c_m.$$

The function $f(\mathbf{x})$ is called the **objective** function, $g(\mathbf{x})$ is called an **inequality constraint**, and $h(\mathbf{x})$ is called an **equality constraint**. In the above problem there are k inequality constraints and m equality constraints. In the following we will always assume that f, g and h are C^1 functions, i.e. that they are differentiable and their derivatives are continuous.

Notice that this problem differs from the regular unconstrained optimization problem in that instead of finding the maximum of $f(\mathbf{x})$, we are finding the maximum of $f(\mathbf{x})$ only over the points which satisfy the constraints.

Example: Maximize $f(x) = x^2$ subject to $0 \le x \le 1$.

Solution: We know that $f(x)$ is strictly monotonically increasing over the domain, therefore the maximum (if it exists) must lie at the largest number in the domain. Since we are optimizing over a compact set, the point $x = 1$ is the maximal number in the domain, and therefore it is the maximum.

This problem was easy because we could visualize the graph of $f(x)$ in our minds and see that it was strictly monotonically increasing over the domain. However, we see a method to find constrained maxima of functions even when we can't picture them in our minds.

2 Equality Constraints

2.1 One Constraint

Consider a simple optimization problem with only one constraint:

$$\max_{x \in \mathbb{R}} f(x_1, \ldots, x_n) \text{ subject to :}$$

$$h(x_1, \ldots, x_n) = c.$$

Now draw level sets of the function $f(x_1, \ldots, x_n)$. Since we might not be able to achieve the unconstrained maxima of the function due to our constraint, we seek to find the value of $\mathbf{x}$ which gets us onto the highest level curve of $f(\mathbf{x})$ while remaining on the function $h(\mathbf{x})$. Notice also that the function $h(\mathbf{x})$ will be just tangent to the level curve of $f(\mathbf{x})$.

Call the point which maximizes the optimization problem $\mathbf{x}^*$, (also referred to as the **maximizer**). Since at $\mathbf{x}^*$ the level curve of $f(\mathbf{x})$ is tangent to the curve $g(\mathbf{x})$, it must also be the case that the gradient of $f(\mathbf{x}^*)$ must be in the same direction as the gradient of $h(\mathbf{x}^*)$, or

$$\nabla f(\mathbf{x}^*) = \lambda \nabla h(\mathbf{x}^*),$$

where λ is merely a constant.

Now define the **Lagrangian** the following way:

$$L(x_1, \ldots, x_n, \lambda) = f(x_1, \ldots, x_n) - \lambda \left[h(x_1, \ldots, x_n) - c\right]$$

Then setting the partial derivatives of this function with respect to x equal to zero will yield the first order conditions for a constrained maximum:

$$\nabla f(\mathbf{x}^*) - \lambda \nabla h(\mathbf{x}^*) = 0.$$

Setting the partial derivative with respect to λ equal to zero gives us our original constraint back:

$$h(\mathbf{x}^*) - c = 0.$$

So the first order conditions for this problem are simply $\nabla L(x, \lambda) = 0$

Example Maximize $f(x_1, x_2) = x_1 x_2$ subject to $h(x_1, x_2) \equiv x_1 + 4x_2 = 16$.

Solution: Form the Lagrangian

$$L(x_1, x_2) = x_1 x_2 - \lambda\left(x_1 + 4x_2 - 16\right)$$

The first order conditions are

$$\frac{dL}{dx_1} = x_2 - \lambda = 0$$

$$\frac{dL}{dx_2} = x_1 - 4\lambda = 0$$

$$\frac{dL}{d\lambda} = x_1 + 4x_2 - 16 = 0$$

From the first two equations we have

$$x_1 = 4x_2.$$

Plugging this into the last equation we have that $x_2 = 2$, which in turn implies that $x_1 = 8$, and that $\lambda = 2$. This states that the only candidate for the solution is $(x_1, x_2, \lambda) = (8, 2, 2)$.

Remember that points obtained using this formula may or may not be a maximum or minimum, since the first order conditions are only necessary conditions. They only give us candidate solutions.

There is another more subtle way that this process may fail, however. Consider the case where $\nabla h(\mathbf{x}^*) = 0$, or in other words, the point which maximizes $f(\mathbf{x})$ is also a critical point of $h(\mathbf{x})$. Remember our necessary condition for a maximum

$$\nabla f(\mathbf{x}^*) = \lambda \nabla h(\mathbf{x}^*),$$

Since $\nabla h(\mathbf{x}^*) = 0$, this implies that $\nabla f(\mathbf{x}^*) = 0$. However, this the necessary condition for an unconstrained optimization problem, not a constrained one! In effect, when $\nabla h(\mathbf{x}^*) = 0$, the constraint is no longer taken into account in the problem, and therefore we arrive at the wrong solution.

2.2 Many Constraints

Consider the problem

$$\max_{x_i \in \mathbb{R}} f(x_1, \ldots, x_n) \text{ subject to :}$$

$$h_1(x_1, \ldots, x_n) = c_1, \ldots, h_m(x_1, \ldots, x_n) = c_m.$$

Let's first talk about how the lagrangian approach might fail. As we saw for one constraint, if $\nabla h(\mathbf{x}^*) = 0$, then the constraint drops out of the equation. Now consider the Jacobian matrix, or a vector of the gradients of the different $h_i(\mathbf{x}^*)$.

$$Dh(\mathbf{x}^*) = \begin{pmatrix} \nabla h_1(\mathbf{x}^*) \\ \vdots \\ \nabla h_m(\mathbf{x}^*) \end{pmatrix} = \begin{pmatrix} \frac{dh_1(\mathbf{x}^*)}{dx_1} & \cdots & \frac{dh_1(\mathbf{x}^*)}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{dh_m(\mathbf{x}^*)}{dx_1} & \cdots & \frac{dh_m(\mathbf{x}^*)}{dx_n} \end{pmatrix}$$

Notice that if any of the $\nabla h_i(\mathbf{x}^*)$'s is zero, then that constraint will not be taken into account in the analysis. Also, there will be a row of zeros in the Jacobian, and therefore the Jacobian will not be full rank. The generalization of the condition that $\nabla h(\mathbf{x}^*) \neq 0$ for the case when $m = 1$ is that the Jacobian matrix must be of rank m. Otherwise, one of the constraints is not being taken into account, and the analysis fails. We call this condition the **non-degenerate constraint qualification (NDCQ).**

Note that we only have to check whether the NDCQ holds at points in the constraint set, since points outside the constraint set are not solution candidates anyways. If we test for NDCQ and find that the constraint is violated for some point within our constraint set, we have to add this point to our candidate solution set. The Lagrangian technique simply does not give us any information about this point.

The Lagrangian for the multi-constraint optimization problem is

$$L(x_1, \ldots, x_n, \lambda) = f(x_1, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i \left[h_i(x_1, \ldots, x_n) - c_i\right]$$

And therefore the necessary conditions for a maximum are

$$\frac{dL}{dx_1} = 0, \ldots, \frac{dL}{dx_n} = 0$$

$$\frac{dL}{d\lambda_1} = 0, \ldots, \frac{dL}{d\lambda_n} = 0.$$

.

Example Maximize $f(x, y, z) = xyz$ subject to $h_1(x, y, z) \equiv x^2 + y^2 = 1$ and $h_2(x, y, z) \equiv x + z = 1$.

First let us form the Jacobian matrix

$$D\mathbf{h}(x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Notice that the rank is 1 if and only if both $x = y = 0$. However, if this is the case, then our first constraint fails to hold. Therefore, the rank is 2 for all points in the constraint set, and so we don't

need to worry about the NDCQ.
Form the Lagrangean

$$L(x, y, z) = xyz - \lambda_1 \left(x^2 + y^2 - 1\right) - \lambda_2 \left(x + z - 1\right).$$

The first order conditions are

$$\frac{\partial L}{\partial x} = yz - 2\lambda_1 x - \lambda_2 = 0$$

$$\frac{\partial L}{\partial y} = xz - 2\lambda_1 y = 0$$

$$\frac{\partial L}{\partial z} = xy - \lambda_2 = 0$$

$$\frac{\partial L}{\partial \lambda_1} = x^2 + y^2 - 1 = 0$$

$$\frac{\partial L}{\partial \lambda_2} = x + z - 1 = 0$$

In order to solve for λ_1 and λ_2 we divide by y, so we have to assume $y \neq 0$. We will treat the case $y = 0$ later. Assuming $y \neq 0$,

$$\lambda_1 = \frac{xz}{2y}, \ \lambda_2 = xy.$$

Plug these into the first equation to get

$$yz - 2\frac{x^2 z}{2y} - xy = 0$$

Multiply both sides by $y \neq 0$

$$y^2 z - x^2 z - xy^2 = 0$$

Now we want to solve the two constraints for y and z in terms of x, plug them into this equation, and get one equation in terms of x

$$(1 - x^2)(1 - x) - x^2(1 - x) - x(1 - x^2) = 0$$

$$(1 - x)\left[(1 - x^2) - x^2 - x(1 + x)\right] = 0$$

$$(1 - x)\left[1 - 3x^2 - x\right] = 0$$

This yields $x = \left\{1, \frac{-1+\sqrt{13}}{6}, \frac{-1-\sqrt{13}}{6}\right\}$.
Let's analyze $x = 1$ first. From the second constraint we have that $z = 0$, and from the first constraint we have that $y = 0$. That contradicts our assumption that $y \neq 0$, so this cannot be a solution.
Plugging in the other values, we get four candidate solutions

$$x = .4343, \ y = \pm 0.9008, \ z = 0.5657$$

$$x = -.7676, \ y = \pm 0.6409, \ z = 1.7676$$

Finally, let's look at the case $y = 0$. Then either $x = 1$ and $z = 0$ or $x = -1$ and $z = 2$, from the equality constraints. In the first case, all the other first order conditions hold as well, so $(1, 0, 0)$ is another candidate solution. In the second case, we get $2 = 0$ in the second FOC and therefore this point cannot be a candidate solution.

3 Inequality Constraints

3.1 One Constraint

Consider the problem

$$\max_{x_i \in \mathbb{R}} f(x_1, \ldots, x_n) \text{ subject to :}$$

$$g(x_1, \ldots, x_n) \leq b.$$

In this case the solution is not *constrained* to the curve $g(\mathbf{x})$, but merely *bounded* by it.

In order to understand the new conditions, imagine the graph of the level sets which we talked about before. Instead of being constrained to the function $g(\mathbf{x})$, the domain is now bounded by it instead. However, the *boundary* of the function is still the same as before. Notice that there is still a point where the boundary is tangent to some level set of $g(\mathbf{x})$. The question now is whether the boundary is *binding* or *not binding*.

Remember that we are trying to find candidate points for a global maximum. Restricting our original domain X to the set of points where $g(x) \leq b$ gives us two types of candidate points: Points on the boundary $g(x) = b$ where $g(x) = b$ is tangent to the level curves of $f(x)$, and local maxima of $f(x)$ for which $g(x) \leq b$. The first type we can find by the constrained FOC $\nabla f(x) = \lambda \nabla g(x)$, and the second type we can find by the unconstrained FOC $\nabla f(x) = 0$. Let's look at each of these in turn.

Case 1: Candidates along the boundary (constraint binding) This is the case where an unconstrained maximum lies outside of the constraint set. In other words, the inequality constrains us from reaching a maximum of f. In this case, the gradient of $f(\mathbf{x})$ is going to point in the steepest direction up the graph. The gradient of $g(\mathbf{x})$ points to the set $g(\mathbf{x}) \geq b$ (since it points in the direction of increasing $g(\mathbf{x})$). Therefore $\nabla g(\mathbf{x})$ is pointing in the same direction as $\nabla f(\mathbf{x})$, which implies that $\lambda \geq 0$. So necessary conditions for a solution on the boundary are

$$\nabla f(x) - \lambda \nabla g(x) = 0 \text{ and } \lambda \geq 0.$$

Case 2: Candidates not on the boundary (constraint not binding) This is the case where an unconstrained maximum lies inside the constraint set. In other words, the inequality does not constrain us from reaching this maximum of f. The first order condition is simply $\nabla f(x) = 0$, which we can rewrite (to take the same shape as above) as

$$\nabla f(x) - \lambda \nabla g(x) = 0 \text{ and } \lambda = 0.$$

In summary, either the constraint is binding (tight), that is $g(x) - b = 0$ and $\lambda \geq 0$, or the constraint is not-binding (slack), and then $\lambda = 0$. We can summarize this new set of conditions as what is called the **complementary slackness conditions**

$$[g(x) - b]\, \lambda = 0.$$

This works because if the constraint is binding, then $g(x) - b = 0$, and if the constraint is not binding, then we want to ignore it by having $\lambda = 0$.

As we can see, it does not matter whether the constraint binds or does not bind, the Lagrangian multiplier must always be greater than or equal to 0. Therefore, another new set of conditions are

$$\lambda_i \geq 0 \;\forall\; i.$$

Finally, we have include our original inequality constraint $g(x) \leq b$.

We can summarize all these FOC in terms of the Lagrangian, which we define as before to be $L(x, \lambda) = f(x) - \lambda(g(x) - b)$:

$$\frac{\partial L(x,\lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} - \lambda \frac{\partial g(x)}{\partial x_i} = 0 \text{ for all} i = 1, \dots, n$$

$$\lambda \frac{\partial L(x,\lambda)}{\partial \lambda} = \lambda \left[g(x) - b \right] = 0 \text{ complementary slackness}$$

$$\frac{\partial L(x,\lambda)}{\partial \lambda} = \left[g(x) - b \right] \leq 0 \text{ original constraint}$$

$$\lambda \geq 0$$

3.2 Many Constraints

Consider the problem:

$$\max_{x_i \in \mathbb{R}} f(x_1, \dots, x_n) \text{ subject to :}$$

$$g_1(x_1, \dots, x_n) \leq b_1, \dots, g_k(x_1, \dots, x_n) \leq b_k.$$

In order to understand the new NDCQ, we must realize that if a constraint does not bind, we don't care if it drops out of the equation. The point of the NDCQ was to ensure that binding constraints do not drop out. Therefore, the NDCQ with inequality constraints is the same as the equality constraints, except for the fact that we only care about the Jacobian matrix of the binding constraints, or the Jacobian for the constraints with $\lambda_i > 0$. (Notice that we cannot tell in advance which constraints will be binding, so we need to check all of them when we check the NDCQ before computing the solution candidates.)

The following first order conditions will tell us the candidate points for a maximum

$$\frac{\partial L}{\partial x_1} = 0, \dots, \frac{\partial L}{\partial x_n} = 0$$

$$\left[g_1(x_1, \dots, x_n) - b_1 \right] \lambda_1 = 0, \dots, \left[g_k(x_1, \dots, x_n) - b_k \right] \lambda_k = 0$$

$$g_1(x_1, \dots, x_n) \leq b_1, \dots, g_k(x_1, \dots, x_n) \leq b_k$$

$$\lambda_1 \geq 0, \dots, \lambda_k \geq 0$$

4 Mixed constraints

Now consider the general problem, in which we have equality and inequality constraints:

$$\max_{x_i \in \mathbb{R}} f(x_1, \dots, x_n) \text{ subject to :}$$

$$g_1(x_1, \dots, x_n) \leq b_1, \dots, g_k(x_1, \dots, x_n) \leq b_k,$$

$$h_1(x_1, \dots, x_n) = c_1, \dots, h_m(x_1, \dots, x_n) = c_m.$$

Summarizing the conditions for equality and inequality constraints that we found above we can formulate the following theorem:

Theorem: Suppose that x^* is a local maximizer of f on the constraint set given by the k inequality and m equalities above. Without loss of generality, assume that the first k_0 inequality constraints are binding and that the other $k - k_0$ inequality constraints are not binding. Further suppose that the Jacobian of the equality constraints and the binding inequality constraints at x^* has full rank. Form the Lagrangian

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_m) = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i [g_i(\mathbf{x}) - b_i] - \sum_{i=1}^{m} \mu_i [h_i(\mathbf{x}) - c_i]$$

Then there exist multipliers $\lambda_1^*, \ldots, \lambda_k^*$ and $\mu_1^*, \ldots, \mu_m^*$ such that

1. $\frac{\partial L}{\partial x_i}(\mathbf{x}^*, \lambda^*, \mu^*) = 0$ for all $i \in \{1, \ldots, n\}$
2. $\lambda_i^*[g_i(\mathbf{x}) - b_i] = 0$ for all $i \in \{1, \ldots, k\}$
3. $h_i(\mathbf{x}^*) = c_i$ for all $i \in \{1, \ldots, m\}$
4. $g_i(\mathbf{x}^*) \leq b_i$ for all $i \in \{1, \ldots, k\}$
5. $\lambda_i \geq 0$ for all $i \in \{1, \ldots, k\}$

Note again that this theorem gives us merely first order, necessary conditions: *If* $\mathbf{x}^*$ is a maximum and *if* the NDCQ holds at $\mathbf{x}^*$, *then* there exist Lagrange multipliers for which the conditions hold true. Finding all tuples $(\mathbf{x}, \lambda, \mu)$ for which the conditions hold will therefore gives us a set of candidate solutions. We still need to check whether these candidates are actually maximizers. (Conditions that can be used to do so will be taught in your first semester math class, or you can read Simon & Blume, chapter 19). Notice also that we may not find all candidate solutions using the Lagrangian method if the NDCQ is violated at some points in the constraint set.

Meaning of the multiplier: See Simon & Blume, Chapter 19.1.

5 Minimization Problems

So far we have only discussed maximization problems. What happens if we are instead looking to minimize a function given certain constraints? There are different ways in which we can change the above technique of using the Lagrangian to apply it to minimization problems. The easiest is probably the following:

Suppose we are given a minimization problem

$$\min_{x_i \in \mathbb{R}} f(x_1, \ldots, x_n) \text{ subject to :}$$

$$g_1(x_1, \ldots, x_n) \leq b_1, \ldots, g_k(x_1, \ldots, x_n) \leq b_k,$$
$$h_1(x_1, \ldots, x_n) = c_1, \ldots, h_k(x_1, \ldots, x_n) = c_k.$$

Finding the minimum of f on a certain domain is really the same as finding the maximum of $-f$ on that domain. So we can transform the above problem into the maximization problem

$$\max_{x_i \in \mathbb{R}} -f(x_1, \ldots, x_n) \text{ subject to :}$$

$$g_1(x_1, \ldots, x_n) \leq b_1, \ldots, g_k(x_1, \ldots, x_n) \leq b_k,$$
$$h_1(x_1, \ldots, x_n) = c_1, \ldots, h_k(x_1, \ldots, x_n) = c_k.$$

We can find candidate solutions for this problem as discussed above. They will also be candidate solutions of the original minimization problem.

6 Kuhn-Tucker Notation

The most common problems in economics are maximization problems dealing with only inequality constraints. Many of these constraints come in the form of non-negativity constraints, such as requiring consumption to be weakly positive. Consider the following problem:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(x_1,\ldots,x_n)$$

$$\text{subject to } g_1(x_1 \le b_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n) \le b_m$$

$$\text{and } x_1 \ge 0,\ldots,x_n \ge 0$$

The Lagrangian function we would write would take the form

$$L(\mathbf{x},\lambda,\mathbf{v}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i\left(g_i(\mathbf{x}) - b_i\right) + \sum_{i=1}^{n} v_i x_i,$$

which would lead to the following first order conditions:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_1} + v_1 = 0,\ldots,\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_n} + v_n = 0$$

$$\lambda_1\left[g_1(\mathbf{x}) - b_1\right] = 0,\ldots,\lambda_m\left[g_m(\mathbf{x}) - b_m\right] = 0$$

$$g_1(x_1,\ldots,x_n) \le b_1,\ldots,g_m(x_1,\ldots,x_n) \le b_m$$

$$v_1x_1 = 0,\ldots,v_nx_n = 0, \text{ and } \lambda_i, x_j \ge 0 \;\forall\; i = 1,\ldots,m \text{ and } \forall\; j = 1,\ldots,n$$

That's a lot of conditions! $(3m + 3n)$

Now consider the Lagrangian without the nonnegativity constraints, and call it the Kuhn-Tucker Lagrangian:

$$L(\mathbf{x},\lambda,\mathbf{v}) = \tilde{L}(\mathbf{x},\lambda) + \sum_{i=1}^{n} v_i x_i$$

The first n first order conditions can be rewritten as

$$\frac{\partial L}{\partial x_j} = \frac{\partial \tilde{L}}{\partial x_j} + v_j = 0 \;\forall\; j = 1,\ldots,n,$$

which implies

$$\frac{\partial \tilde{L}}{\partial x_j} = -v_j \;\forall\; j = 1,\ldots,n.$$

Plugging those into the nonnegativity constraints we have that

$$x_j\frac{\partial \tilde{L}}{\partial x_j} = 0 \text{ and } \frac{\partial \tilde{L}}{\partial x_j} \le 0.$$

Also notice that

$$\frac{\partial L}{\partial \lambda_j} = \frac{\partial \tilde{L}}{\partial \lambda_j} = b_j - g_j(\mathbf{x}) \ge 0 \;\forall\; j = 1,\ldots,m,$$

which implies

$$\frac{\partial \tilde{L}}{\partial \lambda_j} \ge 0 \text{ and } \lambda_j\frac{\partial \tilde{L}}{\partial \lambda_j} = 0 \;\forall\; j = 1,\ldots,m.$$

In summary, the following are the first order conditions for the Kuhn-Tucker Lagrangian:

$$\frac{\partial \tilde{L}}{\partial x_j} \leq 0 \text{ and } x_j \frac{\partial \tilde{L}}{\partial x_j} = 0 \ \forall \ j = 1, \ldots, n$$

$$\frac{\partial \tilde{L}}{\partial \lambda_j} \geq 0 \text{ and } \lambda_j \frac{\partial \tilde{L}}{\partial \lambda_j} = 0 \ \forall \ j = 1, \ldots, m$$

$$\lambda_j \geq 0 \forall j = 1, \ldots m$$

This is $2n + 3m$ constraints, n less than before. If we are dealing with a lot of non-negativity constraints it is therefore faster to use the Kuhn-Tucker Lagrangian.

Exercises

1. There are two commodities: x and y. Let the consumer's consumption set be $\mathbb{R}^2_+$ and his preference relation on his consumption set be represented by $u(x, y) = -(x-4)^2 - y^2$. When his initial wealth is 2 and the relative price is 1, solve his utility maximization problem if it is well defined.
2. Let $f : \mathbb{R}_+ \to \mathbb{R}$ and $f(x) = -(x+1)^2 + 2$. Solve the maximization problem if it is well defined.
3. Let $f : \mathbb{R}^2_+ \to \mathbb{R}$ and $f(x, y) = 2y - x^2$. When (x, y) must be on the unit disc, i.e., $x^2 + y^2 \leq 1$, solve the *minimization* problem if it is well defined.[1]

7 Homework

1. In chapter 18 of Simon and Blume, numbers 3, 6, 7, 10, 11, and 12.

[1]This is the same problem as in Example 18.11 of Simon and Blume (1994).