Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Constrained optimization

Optimization with inequality constraints

The most important class of optimization problems in economics considers maximizing (or minimizing) an objective function subject to k inequality constraints. In these problems, the feasible set takes the form

$$F = \{\boldsymbol{x} \in \mathbb{R}^n : h(\boldsymbol{x}) \leq 0\},$$

where $h(\boldsymbol{x}) \leq 0$ can written more fully as:

$$\begin{array}{rcl} h_1(x_1, ..., x_n) & \leq & 0 \\ \vdots & & \vdots \\ h_k(x_1, ..., x_n) & \leq & 0 \end{array}$$

Notice that we can incorporate equality constrains into these problems since $\{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) = 0\}$ is the same set as $\{\boldsymbol{x} \in \mathbb{R}^n : g(x) \leq 0, \quad -g(\boldsymbol{x}) \leq 0\}$.

Kuhn-Tucker or Karush-Kuhn-Tucker first-order conditions

We shall concentrate on the first-order conditions for an optimum. From this point on, we devote most of our attention to i) maximization problems where the objective function is quasiconcave and ii) minimization problems where the objective function is quasiconvex. For each of these cases, we assume that the constraint functions h_j are quasiconvex so that the feasible set that is given as the intersection of lower level sets of these functions is convex.

This restriction on the constraint functions means that the only types of equality constraints that are allowed are affine linear constraints (since the only functions of $n > 1$ variables that are both quasiconvex and quasiconcave are affine linear functions (i.e. linear plus a constant vector).

Under these conditions, any point satisfying the first-order conditions is a global optimum if the derivative of the objective function at the point in question is non-zero.

We say that an inequality constraint $h_j(x_1, ..., x_n) \leq 0$ is binding at $\hat{\boldsymbol{x}}$ if $h_j(\hat{\boldsymbol{x}}) = 0$. If $h_j(\hat{\boldsymbol{x}}) < 0$, then we say that the constraint is not binding. A non-binding constraint does not restrict the feasible directions for small changes in $\hat{\boldsymbol{x}}$. For binding constraints $h_j(\hat{\boldsymbol{x}})$, the feasible directions $\Delta \boldsymbol{x}$ are given again by:

$$Dh_j(\hat{\boldsymbol{x}})\Delta \boldsymbol{x} \leq 0.$$

Hence the binding constraints are similar to the equality constraints that we discussed in the previous section. The fact that we allow for negative local changes in the values of the constraint functions with inequalities allows us to determine the sign of the Lagrange multipliers (in contrast to the equality constrained case). Non-binding constraints can be ignored. The problem in general is that we do not know a priori which constraints are binding and which are not.

Let's write the Lagrangean function for the optimization problem as before:

$$\mathcal{L}(\boldsymbol{x}, \lambda_1, ..., \lambda_k) = f(\boldsymbol{x}) - \sum_{j=1}^{k} \lambda_j h_j(\boldsymbol{x}).$$

I have adopted the notation from the textbook to denote the Lagrange multipliers in inequality constrained problems by λ_j. If a constraint is not binding, it can be ignored in the problem. If it binds, then it cannot be ignored. But both of these cases are incorporated in the following complementary slackness condition. For all j, we have:

$$\lambda_j h_j(\hat{x}_1, ..., \hat{x}_n) = 0.$$

This simply says that if $h_j(\hat{x}) < 0$, then $\lambda_j = 0$ and the constraint vanishes from the Lagrangean. If the constraint binds, then $h_j(\hat{\boldsymbol{x}}) = 0$ and the complementary slackness is also satisfied.

Based on these considerations, we formulate the first order conditions for $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$ as follows. We consider a point where the constraint qualification holds (i.e. the derivatives of the binding constraints are linearly independent so that we can use implicit function theorem). [1]

The first-order conditions for the problem also known as the Kuhn-Tucker or Karush-Kuhn-Tucker conditions for the problem are given by:

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) = \frac{\partial f}{\partial x_i}(\hat{\boldsymbol{x}}) - \sum_{j=1}^{k} \hat{\lambda}_j \frac{\partial h_j}{\partial x_i}(\hat{\boldsymbol{x}}) = 0 \text{ for all } i,$$

$$\hat{\lambda}_j h_j(\hat{\boldsymbol{x}}) = 0 \text{ for all } j \in \{1, ..., k\},$$

$$\hat{\lambda}_j \geq 0 \text{ for all } j \in \{1, ..., k\},$$

$$h_j(\hat{\boldsymbol{x}}) \leq 0 \text{ for all } j \in \{1, ..., k\}.$$

Let me sum up: at the optimal point $\hat{\boldsymbol{x}}$, we need i) the usual first-order condition for the Lagrangean with respect to the choice variables. ii) we need that $\hat{\boldsymbol{x}}$ be feasible, i.e. $h_j(\hat{\boldsymbol{x}}) \leq 0$ for all j, iii) the complementary slackness conditions, and the non-negativity of the multipliers.

We have not discussed the non-negativity of the multipliers yet, but it is easy to see why this must be true in the case of a single inequality constraint. Assume constraint qualification, i.e. $Dh(\hat{\boldsymbol{x}}) \neq 0$. By the first order conditions with respect to the x_i, we see that as before,

$$\nabla f(\hat{\boldsymbol{x}}) = \lambda \nabla h(\hat{\boldsymbol{x}}).$$

If the multiplier was strictly negative at an optimal point $\hat{\boldsymbol{x}}$, where the constraint binds, then

$$Dh(\hat{\boldsymbol{x}}) \nabla f(\hat{\boldsymbol{x}}) = \lambda \nabla h(\hat{\boldsymbol{x}}) \cdot \nabla h(\hat{\boldsymbol{x}}) \leq 0.$$

Hence movement in the direction of the fastest increase of f is feasible and $\hat{\boldsymbol{x}}$ cannot be an optimum unless $\nabla f(\hat{\boldsymbol{x}}) = 0$. But in this case, $\lambda = 0$ since $\nabla h(\hat{\boldsymbol{x}}) \neq 0$ by constraint qualification contradicting the assumption that $\lambda < 0$.

The general case for the positive sign of the multipliers is proved using either separating hyperplane theorem or Farkas' Lemma and it is left for future studies. The following pictures should give you an idea why the gradient of the objective function must be a positive combination of the gradients of the constraint functions.

[1] I note here that for the case of quasiconvex constraint functions, a sufficient condition for constraint qualification is that the feasible set has an interior point.

Figure 1: Single inequality constraint

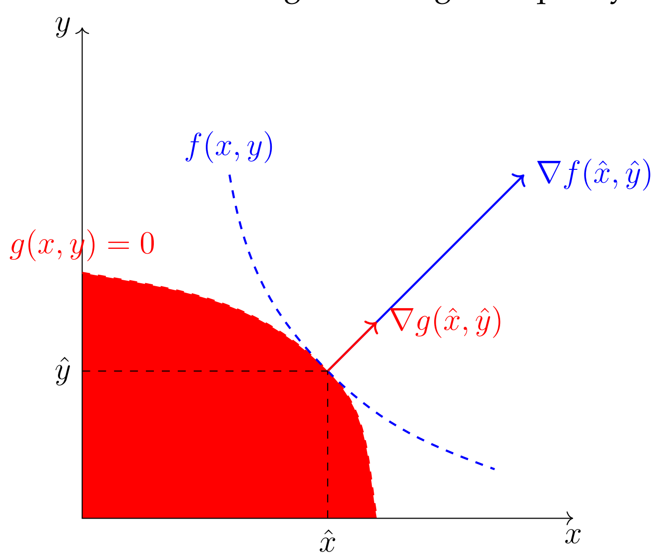
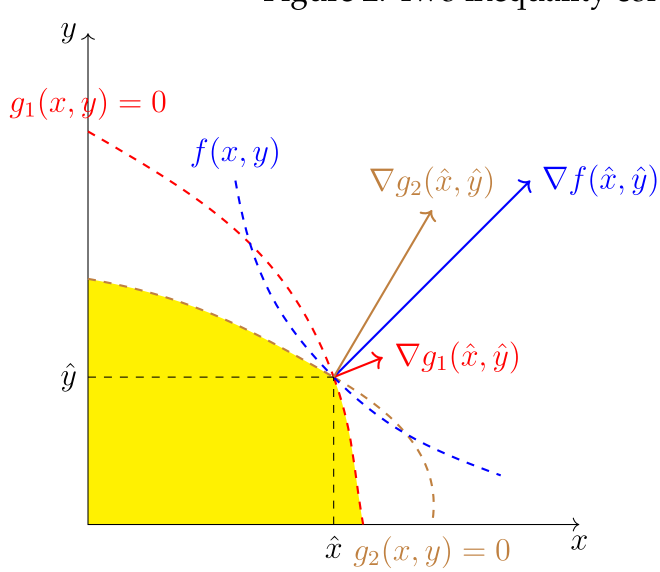


Figure 2: Two inequality constraints



Concave programming

We are now ready to see why the first-order conditions are sufficient for maxima of quasiconcave functions with a non-vanishing derivative on a convex set. Recall from Lecture 6 that a differentiable function f on a convex set X is quasiconcave if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in X$:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) \Rightarrow Df(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) \geq 0.$$

This implies the following (almost converse) result:

Proposition 1. Suppose $Df(\boldsymbol{x})$ is non-zero for all $\boldsymbol{x} \in X$ and f is quasiconcave on X. Then $\hat{\boldsymbol{x}}$ is a global maximum for f on X if $Df(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) \leq 0$ for all $\boldsymbol{y} \in X$

Theorem 1. Suppose that f is quasiconcave and $Df(\boldsymbol{x}) \neq 0$ on a the convex set $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0$ for $j \in \{1, ..., k\}$, where each $h_j(\boldsymbol{x})$ is a quasiconvex function. Then any point satisfying the first-order conditions is a global maximum for f on X.

Proof. Write the first-order condition with respect to $\boldsymbol{x}$ as:

$$Df(\hat{\boldsymbol{x}}) - \sum_{j=1}^{k} \hat{\lambda}_j Dh_j(\hat{\boldsymbol{x}}) = 0. \tag{1}$$

Multiply on the right by $(\boldsymbol{y} - \hat{\boldsymbol{x}})$ to get

$$Df(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) - \sum_{j=1}^{k} \hat{\lambda}_j Dh_j(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) = 0. \tag{2}$$

For feasible directions for binding constraints, we have $Dh_j(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) \leq 0$ since each h_j is assumed to be quasiconvex. For nonbinding constraints, $\hat{\lambda}_j = 0$. Therefore since $\hat{\lambda}_j \geq 0$ for all j, we have

$$\hat{\lambda}_j Dh_j(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) \leq 0 \text{ for all } j.$$

Thus by equation (2), we see that

$$Df(\hat{\boldsymbol{x}})(\boldsymbol{y} - \hat{\boldsymbol{x}}) \leq 0$$

for all feasible $\boldsymbol{y}$. Therefore by the proposition above, $f(\hat{\boldsymbol{x}}) \geq f(\boldsymbol{y})$ for all feasible $\boldsymbol{y}$. □

Before getting into economic applications proper, let's conclude this section with a couple of numerical examples demonstrating how to find constrained maxima.

Example 1. Maximize the objective function $f(x, y, z) = xyz + z$, subject to

$$\begin{aligned} x^2 + y^2 + z &\leq 6 \\ x &\geq 0 \\ y &\geq 0 \\ z &\geq 0. \end{aligned}$$

1. Find the points that satisfy the first-order condition
2. Investigate whether the constraint $x^2 + y^2 + z \leq 6$ binds at optimum.
3. Find a point satisfying the first order conditions with $x = 0$.

Let's do this mechanically and think about what we do only afterwards.

1. From the Lagrangean:

$$\mathcal{L}(x, y, z, \lambda_i) = xyz + z - \lambda_1 \left[x^2 + y^2 + z - 6\right] + \lambda_2 x + \lambda_3 y + \lambda_4 z$$

   The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = yz - 2\lambda_1 x + \lambda_2 = 0 \tag{3}$$
$$\frac{\partial \mathcal{L}}{\partial y} = xz - 2\lambda_1 y + \lambda_3 = 0 \tag{4}$$
$$\frac{\partial \mathcal{L}}{\partial z} = xy + 1 - \lambda_1 + \lambda_4 = 0 \tag{5}$$
$$\lambda_1 \left[x^2 + y^2 + z - 6\right] = 0 \tag{6}$$
$$\lambda_2 x = 0 \tag{7}$$
$$\lambda_3 y = 0 \tag{8}$$
$$\lambda_4 z = 0 \tag{9}$$
$$x^2 + y^2 + z \leq 6 \tag{10}$$
$$-x \leq 0 \tag{11}$$
$$-y \leq 0 \tag{12}$$
$$-z \leq 0 \tag{13}$$
$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3, 4\} \tag{14}$$

2. If at optimum, $\lambda_1 = 0$, i.e. the first constraint is not binding, we get from (5):

$$xy + 1 + \lambda_4 = 0$$

This is not possible for any feasible (x, y), since $\lambda_4 \geq 0$, implying that either x or y must be negative. By the non-negativity constraints, we conclude that

$$\lambda_1 > 0,$$

and $x^2 + y^2 + z \leq 6$ binds at the optimum.

3. Find a critical point with $x = 0$:

$$\frac{\partial \mathcal{L}}{\partial x} = yz + \lambda_2 = 0 \tag{15}$$
$$\frac{\partial \mathcal{L}}{\partial y} = -2\lambda_1 y + \lambda_3 = 0 \tag{16}$$
$$\frac{\partial \mathcal{L}}{\partial z} = 1 - \lambda_1 + \lambda_4 = 0 \tag{17}$$
$$\lambda_1 \left[y^2 + z - 6\right] = 0 \tag{18}$$
$$\lambda_2 x = 0 \tag{19}$$
$$\lambda_3 y = 0 \tag{20}$$
$$\lambda_4 z = 0 \tag{21}$$
$$y^2 + z \leq 6 \tag{22}$$
$$-x \leq 0 \tag{23}$$
$$-y \leq 0 \tag{24}$$
$$-z \leq 0 \tag{25}$$
$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3, 4\} \tag{26}$$

By part b), we know that $\lambda_1 > 0$. The second FOC gives $\lambda_3 = \lambda_1 y$. Multiplying both sides by y and using $y\lambda_3 = 0$, and $\lambda_1 > 1$ we see that $y = 0$. This yields $z = 6$. We have therefore found the point $(x = 0, y = 0, z = 6)$ and $(\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 0)$ satisfying the first-order conditions and $x = 0$.

You can check that the constraints are quasiconvex, but unfortunately the objective function cannot be seen to be quasiconcave. Hence we need to check other cases. By symmetry in the problem, consider $\hat{x} = \hat{y} > 0$ and maximize $\hat{x}^2 z + z$ subject to $2\hat{x}^2 + z = 6$. You can solve from the first-order conditions that $(x = 1, y = 1, z = 4)$ and $(\lambda_1 = 2, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 0)$ is another critical point of the Lagrangean. You can see that this gives a higher value to the objective function and you can argue using symmetry that there are no other critical points. Hence you have found the maximum. In fact, finding a point satisfying the first-order conditions tells us that the objective function is not quasiconcave.

The next problem is to find a maximum in a problem that is really a consumer optimization problem without the economics terminology.

Example 2. Maximize

$$f(x,y) = \alpha x + \sqrt{y}$$

subject to

$$\begin{aligned} px + y &\leq 1, \\ x &\geq 0, \\ y &\geq 0. \end{aligned}$$

Let's assume that $p > 0$ and let's find interior solutions, i.e. $(\hat{x}, \hat{y}) > 0$. Are there other kinds of solutions?

Form the Lagrangean:

$$\mathcal{L}(x, y, \lambda_i) = \alpha x + \sqrt{y} - \lambda_1 [px + y - 1] + \lambda_2 x + \lambda_3 y$$

First-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} = \alpha - \lambda_1 p + \lambda_2 = 0 \quad (27)$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{1}{2} y^{-\frac{1}{2}} - \lambda_1 + \lambda_3 = 0 \quad (28)$$

$$\lambda_1 [px + y - 1] = 0 \quad (29)$$

$$\lambda_2 x = 0 \quad (30)$$

$$\lambda_3 y = 0 \quad (31)$$

$$px + y \leq 1 \quad (32)$$

$$-x \leq 0 \quad (33)$$

$$-y \leq 0 \quad (34)$$

$$\lambda_i \geq 0 \;\; i \in \{1, 2, 3\} \quad (35)$$

First-order conditions give:

$$\lambda_1 = \frac{\alpha + \lambda_2}{p} > 0$$

If $x, y > 0$, we have $\lambda_2, \lambda_3 = 0$ and $y = (\frac{p}{2\alpha})^2$.

Since $\lambda_1 > 0$, we have $px + y = 1$ at optimum and:

$$x^* = \frac{4\alpha^2 - p^2}{4p\alpha^2}$$

Note that this solution is valid only if $2\alpha \geq p$. Constraint qualification holds since the derivative of the binding constraint is nonzero at optimum: $D(h_1(x, y)) = (p, 1)^\top$. If $p > 2\alpha$, then the optimum is a corner solution. From (28), we see that at any optimum, $y > 0$. Therefore the only other possibility is that $(\hat{x}, \hat{y})$ = (0,1). What is the value of λ_2 in this case?