Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Constrained optimization

Motivating examples

Maximum on a closed interval

Consider finding the maximum for $f(x) = 3 + 2x - x^2$ on the feasible set $F = \{x : -\infty < a \leq x \leq b < \infty$. Since f is continuous and the feasible set F is compact, Weierstrass' theorem (see notes on Analysis for this) guarantees the existence of a maximizer, i.e. an $x \in F$ such that for all $y \in F$, we have $f(x) \geq f(y)$.

Notice that f is strictly increasing for $x < 1$ (since $f'(x) > 0$) and strictly decreasing for $x > 1$ since $f'(x) < 0$. If $a \leq 1 \leq b$, then the function is maximized at its critical point $x = 1$. We say that a direction h is feasible from $x_0 \in F$ if $x_0 + h \in F$ for h small enough. Linear approximation by the derivative gives:

$$f(x_0 + h) - f(x_0) = f'(x_0)h.$$

If we have a maximum at x_0, then for all feasible direction

$$f'(x_0)h \leq 0.$$

If $a < x_0 < b$, then we must have $f'(x_0) = 0$ since both directions $h > 0$ and $h < 0$ are feasible. If $f'(x_0) > 0$, then $h > 0$ cannot be feasible if x_0 is a maximum. Therefore $x_0 = b$ if x_0 is the optimal choice and $f'(x_0) > 0$. Similarly, if $f'(x_0) < 0$ and x_0 is the optimum, then $x_0 = a$.

If all directions are feasible from x_0 and x_0 is a maximum, then just as in the case of unconstrained optimization, we must have $f'(x_0)$. For the other cases, the derivative of the objective function at optimum is closely related to the constraint that binds (i.e. restricts the feasible directions).

Utility maximization

A consumer chooses how to spend her wealth w on two goods x_1, x_2 whose prices are p_1, p_2. If the consumer has a continuously differentiable utility function $u(x_1, x_2)$ that has strictly positive marginal utilities in the consumption of each of the goods, then the utility function has no critical points. The feasible set (budget set) $\{(x_1, x_2) | x_1 \geq 0, x_2 \geq 0, p_1 x_1 + p_2 x_2 \leq w\}$ is closed (since it is defined by weak inequalities) and bounded (since $x_i \leq \frac{w}{\min\{p_1, p_2\}}$). By Weierstrass theorem, a utility maximizing choice exists. What can we say about it?

It is not possible to have the optimal consumption at a point $(\hat{x}_1, \hat{x}_2)$, where $p_1 \hat{x}_1 + p_2 \hat{x}_2 < w$. To see this, note that for any such point, $(\hat{x}_1 + h, \hat{x}_2) \in F$ for $h > 0$ small enough. Since u is strictly increasing in both consumptions,

$$u(\hat{x}_1 + h, \hat{x}_2) > u(\hat{x}_1, \hat{x}_2).$$

Hence we know that the optimum satisfies the budget constraint with equality: $p_1 \hat{x}_1 + p_2 \hat{x}_2 = w$.

By the implicit function theorem, you can draw an indifference curve through any point on the budget line. If the budget line and the indifference curve intersect, some points on the budget line are better than the intersecting point (they are on the better side of the indifference curve). Hence such an intersection point cannot be an optimum. Hence the only possibilities for an optimum are: i) a point of tangency between the budget line and the indifference curve and ii) a solution where $x_1 = 0$ or $x_2 = 0$. The first case is called an interior solution and the latter is called a corner solution.

Our task is to come up with a general framework for finding optimum points for general (differentiable) objective functions $f : F \to \mathbb{R}$, where $F \subset \mathbb{R}^n$. With more than two dimensions in the domain, graphical arguments are not possible and we need a general method to handle the constraints.

Optimization with a single equality constraint

We start with local considerations. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function to be maximized and suppose the constraints take the form $g(\boldsymbol{x}) = g(x_1, ..., x_n) = 0$, where $g(\cdot)$ is also assumed to be differentiable. In other words, $F = \{\boldsymbol{x} : g(\boldsymbol{x}) = 0\}$. We write the maximization problem often as:

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{subject to } g(\boldsymbol{x}) = 0.$$

A solution to this problem finds a point $\hat{\boldsymbol{x}}$ such that $f(\hat{\boldsymbol{x}}) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in F$. What can we say about such an $\hat{\boldsymbol{x}}$? If it exists,then for small Δ,

$$f(\hat{\boldsymbol{x}} + \Delta(\boldsymbol{x} - \hat{\boldsymbol{x}})) - f(\hat{\boldsymbol{x}}) = Df(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})\Delta \leq 0$$

for directions $(\boldsymbol{x} - \hat{\boldsymbol{x}})$ such that the constraint is satisfied at $\hat{\boldsymbol{x}} + \Delta(\boldsymbol{x} - \hat{\boldsymbol{x}})$. But how do we know which directions are feasible?

To find the feasible directions, we go back to implicit function theorem. If $\hat{\boldsymbol{x}} \in F$ and $\frac{\partial g}{\partial x_i}(\hat{\boldsymbol{x}}) \neq 0$ for some $i \in \{1, ..., n\}$, then we can find a function $x_i = h(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) =: h(\boldsymbol{x}_{-i})$ in a neighborhood of $\hat{\boldsymbol{x}}_{-i}$ so that

$$g(h(\boldsymbol{x}_{-i}), x_{-i}) = 0.$$

Notice that it is not possible to use the implicit function theorem if at a critical point of the constraint function. Therefore, we must assume that $Dg(\hat{x}) \neq 0$. We call this the constraint qualification and we will see different versions of this in more complex situations with multiple constraints.

Since the function g is at constant value in the feasible set, we have for all feasible directions $(\boldsymbol{x} - \hat{\boldsymbol{x}})$:

$$\nabla g(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0.$$

Notice also that if $(\boldsymbol{x} - \hat{\boldsymbol{x}})$ is feasible, then also $-(\boldsymbol{x} - \hat{\boldsymbol{x}})$ is feasible. From the linear approximation above, this means immediately that for all feasible directions,

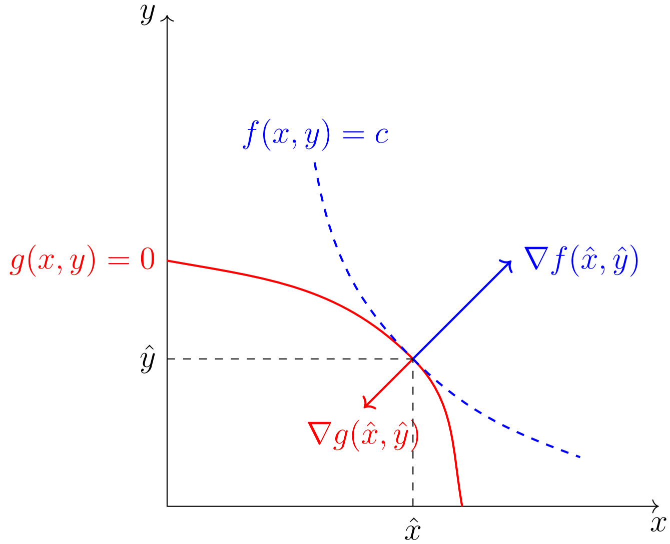
$$Df(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0.$$

But therefore we have shown that at optimum $\hat{\boldsymbol{x}}$,

$$\nabla f(\hat{\boldsymbol{x}}) = \mu \nabla g(\hat{\boldsymbol{x}}).$$

We have derived the following necessary condition for a constrained optimum at $\hat{\boldsymbol{x}}$: the gradient of the objective function must be a scalar multiple of the gradient of the constraint function at the optimum. The second requirement is that the choice must be feasible, i.e. $g(\hat{\boldsymbol{x}}) = 0$.

Figure 1: Single equality constraint



The Lagrangean function

The previous discussion motivates the following function that incorporates the constraints into an augmented objective function called the Lagrange's function or the Lagrangean of the problem.

For a constrained optimization problem with a single equality constraint, we define the following function of $n+1$ variables:

$$\mathcal{L}(\boldsymbol{x}, \mu) = f(\boldsymbol{x}) - \mu g(\boldsymbol{x}).$$

We call the new variable μ the Lagrange multiplier. We will give it a good economic interpretation later in the course. We are interested in the critical points of this augmented function. Therefore we look for $(\hat{\boldsymbol{x}}, \hat{\mu})$ such that

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{\boldsymbol{x}}, \hat{\mu}) = \frac{\partial f}{\partial x_i}(\hat{\boldsymbol{x}}) - \hat{\mu}\frac{\partial g}{\partial x_i}(\hat{\boldsymbol{x}}) = 0 \text{ for all } i,$$
$$\frac{\partial \mathcal{L}}{\partial \mu}(\hat{\boldsymbol{x}}, \hat{\mu}) = g(\hat{\boldsymbol{x}}) = 0.$$

As argued above, these are the first-order conditions for the constrained optimization problem. In order to know if we have found a local maximum or a minimum, we need to look at the second-order Taylor -approximations and the definiteness of the Hessian matrix at the critical point.

As before, write the second-order Taylor approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\hat{\boldsymbol{x}}$ as:

$$f(x) = f(\hat{\boldsymbol{x}}) + Df(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) + (\boldsymbol{x} - \hat{\boldsymbol{x}}) \cdot Hf(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}).$$

If $\hat{\boldsymbol{x}}$ is a maximum, then for all feasible directions $(\boldsymbol{x} - \hat{\boldsymbol{x}})$, we have:
i) $Df(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0$,
ii) $(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cdot Hf(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) \leq 0$.
Since the feasible directions are give by vectors $(\boldsymbol{x} - \hat{\boldsymbol{x}})$ such that

$$\nabla g(\hat{\boldsymbol{x}}) \cdot (\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0,$$

the condition for having a local maximum at $\hat{\boldsymbol{x}}$ is equivalent to checking the negative definiteness of the bordered Hessian where we need

1. The Lagrangean $\mathcal{L}$,
2. The Equality constraint h.

To get the bordered Hessian, start with the derivative of the Lagrangean with respect to the choice variables $\boldsymbol{x}$ at the critical point $\hat{\boldsymbol{x}}$: $H_{\boldsymbol{x}}\mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})$ and 'border' it with the derivative of the constraint function (to capture the restriction to feasible directions).

$$H\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{\boldsymbol{x}}) \\ [Dg(\hat{\boldsymbol{x}})]^T & H_{\boldsymbol{x}}\mathcal{L}(\hat{x}) \end{bmatrix}$$

In the special case where we have only two choice variables, I let the variables be x, y for notational ease, we need to examine

$$H\mathcal{L} = \begin{bmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

How do we determine the negative definiteness of the bordered Hessian?

1. Leading principal minors must alternate in sign [1]

2. $\det H\mathcal{L}(\hat{\boldsymbol{x}})$ must have the same sign as $(-1)^n$.

How many principal minors to examine?

- You need to check the sign of the last $(n-1)$ leading principal minors
- For completeness, I state here that with more constraints, you need to border the Hessian with the derivatives of all binding constraints. If you have k such constraints, then you need to examine the sign of $(n-k)$ leading principal minors.

Bordered Hessians are a bit of a nightmare for me. They are tedious to compute and they tell nothing of significance in the end. In most cases, we can argue by Weierstrass' theorem that a maximum exists and therefore it has to be at a critical point of the Lagrangean (since this is a necessary condition). If there is a single critical point, it must be the maximum. Otherwise one must compare the values of the objective function at the critical points to find the global maximum. (Sometimes it is not entirely trivial to see how to apply Weierstrass' theorem, but in most cases, this is not difficult.)

In any case, here is finally a concrete example:

Example 1. Find the minima and maxima of $f(x,z) = x + z^2$ subject to

$$x^2 + z^2 = 1$$

Form the Lagrangean

$$\mathcal{L}(x,z,\mu) = x + z^2 - \mu(x^2 + z^2 - 1)$$

Differentiate to get the first-order conditions (FOC):

[1] Recall that a leading principal minor of k^{th} order is obtained from a matrix $\boldsymbol{A}$ by deleting its last k rows and columns.

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - 2\mu x = 0 \tag{1}$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2z - 2\mu z = 0 \tag{2}$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = 1 - x^2 - z^2 = 0 \tag{3}$$

The second FOC gives:

$$z(2 - 2\mu) = 0$$

Therefore either $z = 0$, or $\mu = 1$.

Consider first the possibility that $z = 0$. In that case, (3) implies that $x = \pm 1$. We get two critical points from (1):

$\left(x = 1, z = 0, \mu = \frac{1}{2}\right)$ and $\left(x = -1, z = 0, \mu = -\frac{1}{2}\right)$

If $\mu = 1$, (1) implies that $x = \frac{1}{2}$. By substituting into (3) we get the critical points:

$\left(x = \frac{1}{2}, z = \frac{\sqrt{3}}{2}, \mu = 1\right)$ and $\left(x = \frac{1}{2}, z = -\frac{\sqrt{3}}{2}, \mu = 1\right)$

As a result, we have four critical points for the Lagrangean. Draw the constraint set and level curves for the objective function to get a guess of the classification of the critical points.

By examining the bordered Hessian, we see that $\left(x = -1, z = 0, \mu = -\frac{1}{2}\right)$ and $\left(x = 1, z = 0, \mu = \frac{1}{2}\right)$ are local minima, and $\left(x = \frac{1}{2}, z = \pm\frac{\sqrt{3}}{2}, \mu = 1\right)$ are local maxima.

Can you show the existence of a maximum? Which of the local maxima is the global maximum?

Multiple equality constraints

Consider next the case, where we have k equality constraints $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), ..., g_k(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^k$. In this case, we have the problem:

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{subject to } g_1(\boldsymbol{x}) = 0,$$

$$g_2(\boldsymbol{x}) = 0,$$

$$\vdots$$

$$g_k(\boldsymbol{x}) = 0.$$

Form the Lagrangean now with k constraints as a function of $n + k$ variables:

$$\mathcal{L}(\boldsymbol{x}, \mu_1, ..., \mu_k) = f(\boldsymbol{x}) - \sum_{j=1}^{k} \mu_j g_j(\boldsymbol{x}).$$

We can proceed exactly as before to use the linear approximations to characterize the feasible directions from $\hat{\boldsymbol{x}}$ as $\{(\boldsymbol{x} - \hat{\boldsymbol{x}}) : Dg(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})\} = 0$. Since the objective function cannot increase at the optimum in any feasible direction, we have that

$$Df(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0 \text{ whenever } Dg(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0.$$

If $Dg(\hat{\boldsymbol{x}})$ has full rank, then this is equivalent to requiring that $Df(\hat{\boldsymbol{x}})$ and $Dg_j(\hat{\boldsymbol{x}})$ must be linearly dependent. Since we assume that $Dg(\hat{\boldsymbol{x}})$ has full rank, this means that there must exist $(\mu_1, ..., \mu_k)$ such that

$$\nabla f(\hat{\boldsymbol{x}}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{\boldsymbol{x}}).$$

Hence we can summarize the three necessary conditions for local maximum:

i) Gradient alignment: $\nabla f(\hat{\boldsymbol{x}}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{\boldsymbol{x}})$,

ii) Constraint holds: $g(\hat{\boldsymbol{x}}) = 0$,

iii) Constraint qualification: $Dg_1(\hat{\boldsymbol{x}}), ..., Dg_k(\hat{\boldsymbol{x}})$ are linearly independent.

The first two can be achieved by requiring that $(\hat{\boldsymbol{x}}, \hat{\mu}_1, ..., \hat{\mu}_k)$ be a critical point of the Lagrangean. The second-order conditions are based on bordered Hessian matrices as explained at the end of the previous subsection.

Let's end this section with another example

Example 2.

Consider the objective function

$$f(x, y, z) = xz + yz$$

and a maximization problem subject to:

$$g_1(x, y, z) = y^2 + z^2 - 1 = 0$$
$$g_2(x, y, z) = xz - 3 = 0$$

There are two ways to approach such a problem: i) Do not look at the problem itself, but just plug in the general Lagrangean machinery. ii) Simplify the problem before forming the Lagrangean. I always opt for the second since for more complicated problems, the first approach becomes nightmarish.

For this particular problem, the obvious simplification seems to be to substitute the second constraint into the objective function. The problem then becomes:

$$\max_{y,z} 3 + yz$$
$$\text{subject to } y^2 + z^2 - 1 = 0.$$

This is obviously much easier to solve and also you can see now that the feasible set $F = \{(y, z) | y^2 + z^2 = 1\}$ is a compact set. Since $f(y, z) = 3 + yz$ is continuous in y, z, Weierstrass' theorem guarantees that a maximum exists and therefore all you need to consider are the critical points of the Lagrangean:

$$\mathcal{L}(y, z, \mu) = 3 + yz - \mu(y^2 + z^2 - 1).$$

The first order conditions for a critical point are:

$$\frac{\partial \mathcal{L}(\hat{y}, \hat{z}, \hat{\mu})}{\partial y} = \hat{z} - 2\hat{\mu}\hat{y} = 0,$$
$$\frac{\partial \mathcal{L}(\hat{y}, \hat{z}, \hat{\mu})}{\partial z} = \hat{y} - 2\hat{\mu}\hat{z} = 0,$$

$$\frac{\partial \mathcal{L}(\hat{y}, \hat{z}, \hat{\mu})}{\partial \mu} = \hat{y}^2 + \hat{z}^2 - 1 = 0.$$

From the first equation, we get: $2\hat{\mu} = \frac{\hat{z}}{\hat{y}}$, and from the second, $2\hat{\mu} = \frac{\hat{y}}{\hat{z}}$. Therefore $\hat{y} = \pm\hat{z}$ and we get from the third: $\hat{y}, \hat{z} = \pm\frac{1}{\sqrt{2}}$.

Therefore $\hat{\mu} = \pm\frac{1}{2}$ and we can solve for the original problem that the following are the critical points:

$$(\hat{x}, \hat{y}, \hat{z}, \hat{\mu}) = (3\sqrt{2}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}),$$

$$(\hat{x}, \hat{y}, \hat{z}, \hat{\mu}) = (3\sqrt{2}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}),$$

$$(\hat{x}, \hat{y}, \hat{z}, \hat{\mu}) = (-3\sqrt{2}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}),$$

$$(\hat{x}, \hat{y}, \hat{z}, \hat{\mu}) = (-3\sqrt{2}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}).$$

For maxima, it is clear that y and z must have the same sign and by the symmetry of the problem you can see that both solutions with $\mu = \frac{1}{2}$ are maxima. You may also want to look graphically at the reduced problem.