Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Examples of constrained optimization (optional)

Optimization with equality constraints

Geometric mean vs. arithmetic mean

One key result in mathematics says that for positive numbers $x_1, ..., x_n > 0$, their arithmetic mean, $AM(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n x_i$ is at least as large as their geometric mean, $GM(\boldsymbol{x}) = \prod_i x_i^{\frac{1}{n}}$. One way to prove this is via constrained optimization.

$$\max_{\boldsymbol{x}\geq 0} x_1^{\frac{1}{n}} x_2^{\frac{1}{n}} \dots x_n^{\frac{1}{n}}$$

$$\text{subject to } \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}.$$

Weierstrass' theorem guarantees that a maximum exists since the feasible set is bounded ($\max_i x_i \leq n\bar{x}$) and closed (determined by an equality and non-negativity constraints), and the objective function is continuous. Since the objective function is at its minimum at $x_i = 0$ for some i, we can ignore the non-negativity constraints since the optimum is always at an interior point $x_i > 0$ for all i.

The Lagrangean for this problem is:

$$\mathcal{L}(\boldsymbol{x}, \mu) = x_1^{\frac{1}{n}} x_2^{\frac{1}{n}} \dots x_n^{\frac{1}{n}} - \mu(\frac{1}{n}\sum_{i=1}^n x_i - \bar{x}).$$

Let $\hat{y} = \hat{x}_1^{\frac{1}{n}} \hat{x}_2^{\frac{1}{n}} \dots \hat{x}_n^{\frac{1}{n}}$. Then we can write the first-order conditions for the critical points of the Lagrangean as:

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})}{\partial x_i} = \frac{1}{n}\frac{\hat{y}}{\hat{x}_i} - \hat{\mu}\frac{1}{n} = 0 \text{ for all } i,$$

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})}{\partial \mu} = \frac{1}{n} \sum_{i=1}^{n} x_i - \bar{x} = 0.$$

The first-order conditions with respect to x_i, x_j imply that

$$\frac{1}{n} \frac{\hat{y}}{\hat{x}_i} = \hat{\mu} \frac{1}{n} = \frac{1}{n} \frac{\hat{y}}{\hat{x}_j},$$

and therefore, $\hat{x}_i = \hat{x}_j$ for all i, j and substituting into the constraint, we get

$$\hat{x}_i = \bar{x} \text{ for all } i.$$

Perhaps the easiest way to see that the critical point is a minimum, is to notice that it is the only point satisfying the necessary conditions for a maximum and Weierstrass' theorem guarantees that a maximum exists. Therefore, the critical point must be the maximum.

We conclude that for all $(x_1, ..., x_n)$ such that $\frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$, we have:

$$GM(\boldsymbol{x}) = x_1^{\frac{1}{n}} x_2^{\frac{1}{n}} \dots x_n^{\frac{1}{n}} \leq \bar{x}^{\frac{1}{n}} \bar{x}^{\frac{1}{n}} \dots \bar{x}^{\frac{1}{n}} = \bar{x} = AM(\boldsymbol{x}).$$

This result generalizes to weighted averages. For $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)$ with $\alpha_i \geq 0$ for all i and $\sum_{i=1}^{n} \alpha_i = 1$, we say that $\boldsymbol{\alpha} \cdot \boldsymbol{x}$ is a weighted average $\bar{x}(\boldsymbol{\alpha})$ of $(x_1, ..., x_n)$. I claim that for all positive $\boldsymbol{x}$,

$$\prod_i x_i^{\alpha_i} \leq \boldsymbol{\alpha} \cdot \boldsymbol{x}.$$

To see this, note first that $\hat{\boldsymbol{x}}$ maximizes $f(\boldsymbol{x})$ in F if and only if it maximizes $g(f(\boldsymbol{x}))$ in F. Therefore consider, the auxiliary maximization problem:

$$\max_{\boldsymbol{x}} \sum_{i=1}^{n} \alpha_i \ln x_i$$

$$\text{subject to } \sum_{i=1}^{n} x_i \alpha_i = \bar{x}(\boldsymbol{\alpha}).$$

The Lagrangean is:

$$\mathcal{L}(\boldsymbol{x}, \mu) = \sum_{i=1}^{n} \alpha_i \ln x_i - \mu(\sum_{i=1} x_i \alpha_i - \bar{x}(\boldsymbol{\alpha})).$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})}{\partial x_i} = \frac{\alpha_i}{\hat{x}_i} - \hat{\mu}\alpha_i = 0 \text{ for all } i,$$
$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})}{\partial \mu} = \sum_{i=1} x_i\alpha_i - \bar{x}(\boldsymbol{\alpha}) = 0.$$

As before, we get that at optimum, $x_i = \bar{x}(\boldsymbol{\alpha})$ for all i and at optimum, the weighted average coincides with the weighted geometric mean. Since $\sum_{i=1}^n \alpha_i \ln x_i$ has a negative definite Hessian matrix, we have proved the claim.

Notice that the maximization problem here is exactly the same as in maximizing a Cobb-Douglas utility function for prices given by the consumption weights α_i in the utility function.

One can also define $f(\boldsymbol{x}; \rho) = (\sum_{i=1}^n x_i\rho)^{\frac{1}{\rho}}$ to be the ρ^{th} power mean of $\boldsymbol{x}$. Note that $f(\boldsymbol{x}, 1) = AM(\boldsymbol{x})$. In Problem Set 1, you were asked to show that $f(\boldsymbol{x}; \rho) \to \prod_i x_i^{\frac{1}{n}} = GM(\boldsymbol{x})$ as $\rho \to 0$. You can show as an exercise using the approach above that for $\rho < 1$, we have: $f(\boldsymbol{x}; \rho) \leq AM(\boldsymbol{x})$ for all $\boldsymbol{x}$, and equality holds only if and only if $(x_i = AM(\boldsymbol{x}) = f(\boldsymbol{x}, 1)$ for all i. Amongst other things, one can also show that $f(\boldsymbol{x}; \rho)$ is increasing in ρ. You can also show that for $\rho > 1$, the maximum value of $f(\boldsymbol{x}; \rho)$ on the feasible set

$$F = \{\boldsymbol{x} \geq 0 \quad | \quad \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}\}$$

is obtained at any corner point $x_i = n\bar{x}$ for some i and $x_j = 0$ for $j \neq i$.

Quadratic optimization and eigenvalues of symmetric matrices

Consider the problem of maximizing the quadratic form $\boldsymbol{x} \cdot A\boldsymbol{x}$ in $\mathbb{R}^n$ (recall that $\boldsymbol{A}$ is required to be symmetric) by choosing a vector $\boldsymbol{x} \in \mathbb{R}^n$ subject to the constraint that $\|\boldsymbol{x}\| = 1$, i.e. that the vector has unit length. Since the feasible set $F = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}$ is a compact set and quadratic forms are continuous, Weierstrass' theorem guarantees that a maximum exists.Write this problem as:

$$\max_{\boldsymbol{x}} \boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x}$$

$$\text{subject to } \boldsymbol{x} \cdot \boldsymbol{x} - 1 = 0.$$

Form the Lagrangean:

$$\mathcal{L}(\boldsymbol{x}, \mu) = \boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} - \mu(\boldsymbol{x} \cdot \boldsymbol{x} - 1).$$

First-order conditions (from quadratic optimization):

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu}) = 2\boldsymbol{A}\hat{\boldsymbol{x}} - 2\hat{\mu}\hat{\boldsymbol{x}} = 0.$$

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{x}}, \hat{\mu})}{\partial \mu} = \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} - 1 = 0.$$

First-order conditions with respect to $\boldsymbol{x}$ show that the solution vector is an eigenvector of A and μ is an eigenvalue. This shows that symmetric matrices have a real eigenvector and a real eigenvalue. Let v^1 denote this eigenvector and μ^1 denote the eigenvalue. We also get that the optimized value is μ^1 since:

$$v^1 \cdot \boldsymbol{A}v^1 = \mu^1 v^1 \cdot v^1 = \mu^1.$$

We want to prove by induction that $\boldsymbol{A}$ has n orthogonal eigenvectors and that all its eigenvalues are real. Suppose then that we have $v^1, ...v^k$ orthogonal eigenvectors and $\mu^1, ..., \mu^k$ real eigenvalues and consider next:

$$\max_{\boldsymbol{x}^{k+1}} \boldsymbol{x}^{k+1} \cdot \boldsymbol{A}\boldsymbol{x}^{k+1}$$

$$\text{subject to } \boldsymbol{x}^{k+1} \cdot \boldsymbol{x}^{k+1} = 1,$$

$$\text{and } \boldsymbol{x}^{k+1} \cdot v^i = 0 \text{ for all } i \in \{1, ..., k\}.$$

Again, a solution exists (by Weierstrass theorem). We want to show that the solution $\boldsymbol{x}^{k+1}$ is an eigenvector of $\boldsymbol{A}$ and the Lagrange multiplier $\mu^{k+1} = \boldsymbol{x}^{k+1}\boldsymbol{A}\boldsymbol{x}^{k+1}$ is the corresponding eigenvalue. The gradients of the constraints $(\boldsymbol{x}^{k+1}, v^1, ..., v^k)$ are linearly independent by construction. Therefore Lagrange's theorem implies that there exist multipliers $\mu^{k+1}, \lambda^1, ..., \lambda^k$ such that

$$2\boldsymbol{A}\boldsymbol{x}^{k+1} - 2\mu^{k+1}\boldsymbol{x}^{k+1} - \sum_{i=1}^{k} \lambda^i v^i = 0. \tag{1}$$

To show the result, we show first that $v^i \cdot \boldsymbol{A}\boldsymbol{x}^{k+1} = 0$ for all $i \in \{1, ..., k\}$. By induction hypothesis, $\boldsymbol{A}v^i = \lambda^i v^i$, so that (since $\boldsymbol{A}$ is symmetric):

$$v^i \cdot \boldsymbol{A}\boldsymbol{x}^{k+1} = \boldsymbol{x}^{k+1} \cdot \boldsymbol{A}v^i = \lambda^i \boldsymbol{x}^{k+1} \cdot v^i = 0.$$

Hence the first two terms in (1) are orthogonal to the terms in the sum. If two orthogonal vectors sum to zero, they must both be zero. To see this, if $\boldsymbol{x}+\boldsymbol{y}=0$ then $(\boldsymbol{x}+\boldsymbol{y})\cdot(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{x}\cdot\boldsymbol{x}+\boldsymbol{y}\cdot\boldsymbol{y}+2\boldsymbol{x}\cdot\boldsymbol{y}=0$, but if $\boldsymbol{x}\cdot\boldsymbol{y}=0$, this implies that $\boldsymbol{x}=\boldsymbol{y}=0$. This means that we must have:

$$2\boldsymbol{A}\boldsymbol{x}^{k+1}=2\mu^{k+1}\boldsymbol{x}^{k+1},$$

and $(\mu^{k+1},\boldsymbol{x}^{k+1})$ is indeed an eigenpair of $\boldsymbol{A}$ as required.

Multiplying (1) from the left by $(\boldsymbol{x}^{k+1})^{\top}$ gives $\boldsymbol{x}^{k+1}\cdot\boldsymbol{A}\boldsymbol{x}^{k+1}=\mu^{k+1}$.