April 27, 2022

3. OPTIMIZATION WITH EQUALITY CONSTRAINTS: LAGRANGE'S METHOD

All throughout this chapter, D denotes an **open subset** of $\mathbb{R}^n$.

1. Introduction

The optimization problme with equality constraints is

$$\max \quad (\text{resp. } \min) \quad f(x) \quad \text{s.a: } x \in S, \tag{1.1}$$

where $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \dots, g_m) : D \longrightarrow \mathbb{R}^m$, with $m < n$ and $S = \{x \in D : g(x) = b\}$, with $b = (b_1, \dots, b_m)$.

1.1. **Elimination of constraints.** If it is possible to find a continuous function $h = (h_1, \dots, h_m)$ that allows to express m variables in terms of the other $n - m$ variables (say the first m variables upon realignment)

$$\begin{aligned} x_1 &= h_1(x_{m+1}, \dots, x_n) \\ &\vdots \\ x_m &= h_m(x_{m+1}, \dots, x_n), \end{aligned}$$

so that this system of equations is equivalent to the original , $g(x) = b$, then $x_0 = (x^0_{m+1}, \dots, x^0_n)$ solves the unconstrained problem

$$\text{opt}\, f(h(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n)$$

if and only if $(h(x_0), x_0)$ solves (1.1).

Example 1.1. Suppose we want to solve

$$\min \; f(x, y, z) = 2x + y - z \quad \text{s.t.: } \left\{ \begin{array}{l} x - y + z = 6, \\ z^2 - y = 0. \end{array} \right.$$

We can solve for x and y in terms of z: $y = z^2$, $x = 6 + z^2 - z$- Plugging these values into the expression of f we have a function of z only

$$F(z) = f(6 + z^2 - z, z^2, z) = 12 + 3z^2 - 3z.$$

The function F in this example is a convex parabola, thus it has a global minimum at the vertex, which is $z^0 = \frac{1}{2}$. Hence the global minimum of the original problem is the point

$$(x^0, y^0, z^0) = \left(\frac{23}{4}, \frac{1}{4}, \frac{1}{2} \right).$$

Of course, to find the function h could be impossible. It is needed to look for other more indirect methods to deal with this kind of problems.

2. Necessary first order conditions: Lagrange Theorem

Let $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \dots, g : m) : D \longrightarrow \mathbb{R}^m$, such that f and g are of class C^1 in D.

Definition 2.1. The point $x_0 \in S$ is a regular point of (1.1) if the rank of the matrix $\mathrm{D}\, g(x_0)$ is m.

Remark 2.2. Intuitively, the non-degenerate constraint qualification means that the set S is a smooth 'surface' in $\mathbb{R}^n$ of dimension $n - m$ and at each point $x \in S$ we can compute the tangent plane $T_x S$ to S as follows

$$x + \{v \in \mathbb{R}^n : \mathrm{D}\, g(x)v = 0\} = \{x + v : v \in \mathbb{R}^n, \quad \mathrm{D}\, g(x)v = 0\} \tag{2.1}$$

Definition 2.3. The Lagrangian function of (1.1) is

$$L(x, \lambda) = f(x) + \lambda \cdot (b - g(x)),$$

where $x = (x_1, \dots, x_n) \in D$, $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$ and

$$\lambda \cdot (b - g(x)) = \sum_{i=1}^{m} \lambda_j (b_i - g_i(x)).$$

Theorem 2.4. If $x_0 \in S$ is a regular point and it is solution of (1.1), then there exists a unique vector $\lambda_0 = (\lambda_1^0, \dots, \lambda_m^0) \in \mathbb{R}^m$ such that (x_0, λ_0) is a critical point of the Lagrangian function of (1.1), that is

$$\nabla_x L(x_0, \lambda_0) = 0,$$
$$\nabla_\lambda L(x_0, \lambda_0) = 0.$$

Definition 2.5. If (x_0, λ_0) is a critical point of the Lagrangian of (1.1), then x_0 is a critical point of f relative to S and that $\lambda_1^0, \dots, \lambda_m^0 \in \mathbb{R}$ are the associated Lagrange multipliers.

Definition 2.6. The equations

$$\nabla_x L(x_0, \lambda_0) = 0,$$
$$\nabla_\lambda L(x_0, \lambda_0) = 0.$$

are the Lagrange equations.

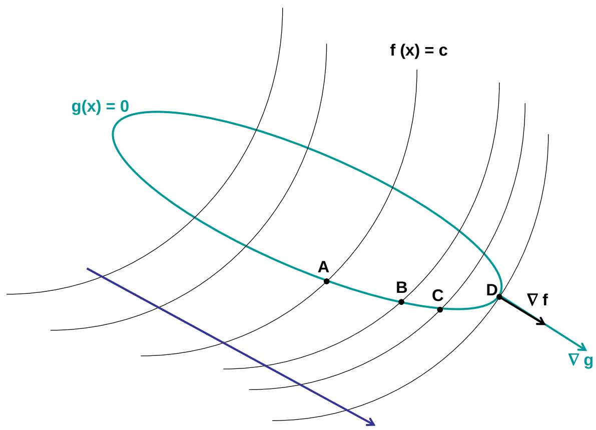
Remark 2.7. There are $n + m$ Lagrange equations and $n + m$ unknowns.
Note that $\nabla_\lambda L(x_0, \lambda_0) = 0$ is the same as $g(x) = b$.

Remark 2.8. If S is compact, then the Theorem of Weierstrass assures the existence of global solutions of (1.1). Assuming that the condition of regularity holds, these global solutions are critical points of f relative to S. We will locate the global solutions by evaluating those critical points with the objective function f. The extremal values will determine the global maximum and minimum of f on S.

Remark 2.9 (Why should the Lagrange equations hold?). To make things simple, let us assume that there is only one restriction and the problem is the following

$$\max f(x)$$
$$\text{s.t.: } g(x) = 0.$$

In the figure we have represented the set $\{x : g(x) = 0\}$ and the level surfaces (curves) $f(x) = c$. The arrow points in the direction in which the function f grows (at each point is given by the gradient of f).



We see that, for example, A cannot be a maximum of f in the set $\{x : g(x) = 0\}$, since f attains a higher value at the point B, that is $f(B) > f(A)$. Likewise, we see graphically that $f(C) > f(B)$. The point D is exactly the point at which if we keep moving in the direction of $\nabla f(D)$ we can no longer satisfy the restriction $g(x) = 0$. At D, the level curves $f(x) = c$ and $g(x) = 0$ are tangent. Equivalently, $\nabla f(D)$ and $\nabla g(D)$ are parallel, so one is a multiple of the other. That is, $\nabla f(D) = \lambda \nabla g(D)$ for some $\lambda \in \mathbb{R}$.

Example 2.10 (Utility Maximization with a budget constraint). Consider the problem

$$\left.\begin{array}{ll} \max & u(x) \\ \text{s.t.} & p \cdot x = t \end{array}\right\}$$

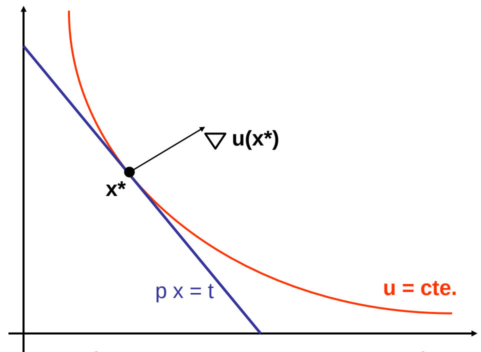
Here we interpret $x \in \mathbb{R}^n_+$ as a consumption bundle and $p \in \mathbb{R}^n_{++}$ as the prices of the goods. The agent has an income t and chooses a consumption bundle x such that it maximizes his utility, subject to the budget constraint, $p \cdot x = t$. The Lagrangian function is

$$L(x, y, \lambda) = u(x) + \lambda(t - p \cdot x)$$

and the Lagrange equations are

$$\left.\begin{array}{l} \nabla u = \lambda p \\ p \cdot x = t \end{array}\right\} \tag{2.2}$$

Thus, if x^* is the bundle that solves the above problem, we have that $\nabla u(x^*)$ is perpendicular to plane $p \cdot x = t$.



On the other hand, we see that equations 2.2 are equivalent to

$$\begin{aligned} \mathrm{MRS}_{ij}(x) &= \frac{p_i}{p_j} \\ px &= t \end{aligned}$$

where

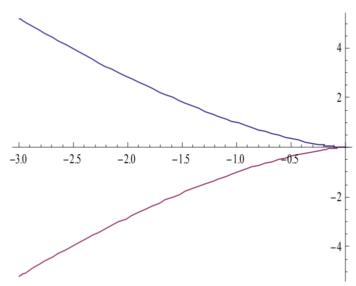
$$\mathrm{MRS}_{ij} = \frac{\partial u \,/\partial x_i}{\partial u \,/\partial x_j}$$

is the marginal rate of substitution between good i and good j.

Example 2.11 (The non-degenerate constraint qualification)*.* This is an example that shows that if the non-degenerate constraint qualification in Definition 2.1 does not hold, then the Lagrange equations might not determine the optimum. Consider the problem

$$\begin{aligned} &\max\ x \\ &\text{s.t.: } x^3 + y^2 = 0. \end{aligned}$$

The set $\{(x, y) \in \mathbb{R}^2 : x^3 + y^2 = 0\}$ is represented in the following figure



Clearly, the solution is $x = 0$. But, if we write the Lagrangian

$$L(x, y, \lambda) = x + \lambda(-x^3 - y^2)$$

the Lagrange equations are

$$\begin{aligned} 1 - 3\lambda x^2 &= 0 \\ -2\lambda y &= 0 \\ x^3 + y^2 &= 0. \end{aligned}$$

The first equation implies that $x \neq 0$ (why?). Using now the third equation we obtain that $y = \sqrt{-x^3} \neq 0$. Since $y \neq 0$ the second equation implies that $\lambda = 0$. But if we plug in $\lambda = 0$ we obtain a contradiction. Therefore the system of the Lagrange equations does not have a solution.
What is wrong? The Jacobian of g is

$$\mathrm{D}\, g(x,y) = (3x^2, 2y)$$

The point $(0,0)$ satisfies the restriction of the problem, but

$$\operatorname{rank}(\mathrm{D}\, g(0,0)) = \operatorname{rank}(0,0) = 0$$

so that the non-degenerate constraint qualification does not hold.

3. Necessary and sufficient second order conditions

Let $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g : m) : D \longrightarrow \mathbb{R}^m$, such that f and g are of class C^2 in D.

Definition 3.1. The Hessian matrix of L with respect to x, $\mathrm{H}_x\, L(x,\lambda)$, is the symmetric matrix which element (i,j) is

$$\frac{\partial^2 L}{\partial x_i \partial x_j}(x,\lambda).$$

Recall the definition of the tangent space 2.1 to the feasible set S at a point $x_0 \in S$

$$x_0 + \{v \in \mathbb{R}^n : \mathrm{D}\, g(x_0)v = 0\}.$$

Definition 3.2. The tangent subspace to the feasible set S at a point $x_0 \in S$ is

$$T_S(x_0) = \{v \in \mathbb{R}^n : \mathrm{D}\, g(x_0)v = 0\}$$

Hence the tangent space can be written $x_0 + T_S(x_0)$.

Theorem 3.3 (Second order necessary conditions)**.** Let x_0 be a regular point of (1.1) that satisfies the first order necessary conditions with associated multiplier λ_0. We have

(1) If x_0 is a local minimum of f on S, then the quadratic form $\mathrm{H}_x\, L(x_0, \lambda_0)$ restricted to $T_S(x_0)$ is positive definite or positive semidefinite.
(2) If x_0 is a local maximum of f on S, then the quadratic form $\mathrm{H}_x\, L(x_0, \lambda_0)$ restricted to $T_S(x_0)$ is negative definite or negative semidefinite.

Corollary 3.4. If the quadratic form $\mathrm{H}_x\, L(x_0, \lambda_0)$ restricted to $T_S(x_0)$ is indefinite, then x_0 is neither a local maximum nor a local minimum of f on S.

Theorem 3.5 (Second order sufficient conditions)**.** Let x_0 be a regular point of (1.1) that satisfies the first order necessary conditions with associated multiplier λ_0. We have

(1) If the quadratic form $\mathrm{H}_x\, L(x_0, \lambda_0)$ restricted to $T_S(x_0)$ is positive definite, thenx_0 is a strict local minimum of f on S.

(2) If the quadratic form $\mathrm{H}_x\, L(x_0, \lambda_0)$ restricted to $T_S(x_0)$ is negative definite, thenx_0 is a strict local maximum of f on S.

Example 3.6. Let us solve the problem

$$\begin{array}{cc} \max & xy \\ \text{s.t.:} & p_1 x + p_2 y = m \end{array}$$

with $m, p_1, p_2 \neq 0$.
Let $f(x,y) = xy$, $g(x,y) = m - p_1 x - p_2 y$. Then, $\nabla g(x,y) = (-p_1, -p_2)$ which does not vanish on the set $S = \{(x,y) \in \mathbb{R}^2 :\ p_1 x + p_2 y = m\}$. Therefore, The regularity condition holds. Let us consider the Lagrangian function

$$L(x,y,\lambda) = xy + \lambda(m - p_1 x - p_2 y).$$

The Lagrange equations are

$$\begin{aligned} y - \lambda p_1 &= 0 \\ x - \lambda p_2 &= 0 \\ p_1 x + p_2 y &= m. \end{aligned}$$

The solution of this system is

$$x_0 = \frac{m}{2p_1}, \qquad y_0 = \frac{m}{2p_2} \qquad \lambda_0 = \frac{m}{2p_1 p_2}.$$

The Hessian matrix of L with respect to (x,y) at the point (x_0, y_0, λ_0) is

$$\mathrm{H}_{(x,y)}\, L(x_0, y_0, \lambda_0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which is non-definite. On the other hand,

$$\begin{aligned} T_S(x_0, y_0) &= \{v \in \mathbb{R}^2 \colon \nabla g\left(\frac{m}{2p_1}, \frac{m}{2p_2}\right) \cdot v = 0\} \\ &= \{(v_1, v_2) \in \mathbb{R}^2 \colon (p_1, p_2) \cdot (v_1, v_2) = 0\} \\ &= \{(t, -\frac{p_1}{p_2} t) \in \mathbb{R}^2 \colon t \in \mathbb{R}\} \end{aligned}$$

Therefore,

$$(t, -\frac{p_1}{p_2} t) \cdot \mathrm{H}\, L_{(x,y)}(x_0, y_0, \lambda_0) \begin{pmatrix} t \\ -\frac{p_1}{p_2} t \end{pmatrix} = (t, -\frac{p_1}{p_2} t) \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} t \\ -\frac{p_1}{p_2} t \end{pmatrix} = -\frac{p_1}{p_2} t^2 < 0$$

if $t \neq 0$. Hence,

$$\left(\frac{m}{2p_1}, \frac{m}{2p_2}\right)$$

is a maximum.

Example 3.7. Let us solve the problem

$$\begin{array}{cc} \max & x^2 + y^2 \\ \text{s.t.:} & xy = 4 \end{array}$$

Let $f(x,y) = x^2 + y^2$, $g(x,y) = xy$. Then $\nabla g(x,y) = 2(y,x)$ which does not vanish on the set $S = \{(x,y) \in \mathbb{R}^2 \colon xy = 4\}$. Therefore, The regularity condition holds. Let us consider the Lagrangian function

$$L(x,y,\lambda) = x^2 + y^2 + \lambda(4 - xy).$$

The Lagrange equations are

$$\begin{aligned} 2x - \lambda y &= 0 \\ 2y - \lambda x &= 0 \\ xy &= 4. \end{aligned}$$

The above system has two solutions

$$x = y = 2,\ \lambda = 2$$
$$x = y = -2,\ \lambda = 2.$$

The Hessian matrix of L with respect to (x, y) at the point $(2, 2, 2)$ is

$$\mathrm{H}_{(x,y)}\, L(2,2,2) = \begin{pmatrix} 2 & -\lambda \\ -\lambda & 2 \end{pmatrix}\Bigg|_{\lambda=2} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix},$$

which is positive semidefinite. On the other hand,

$$\begin{aligned} T_S(2,2) &= \{v \in \mathbb{R}^2 \colon \nabla g(2,2) \cdot v = 0\} \\ &= \{(v_1, v_2) \in \mathbb{R}^2 \colon (2,2) \cdot (v_1, v_2) = 0\} \\ &= \{(t, -t) \in \mathbb{R}^2 \colon t \in \mathbb{R}\} \end{aligned}$$

Hence,

$$(t, -t) \cdot \mathrm{H}_{(x,y)}\, L(2,2,2) \begin{pmatrix} t \\ -t \end{pmatrix} = (t, -t) \cdot \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} t \\ -t \end{pmatrix} = 8t^2$$

so that $\mathrm{H}_{(x,y)}\, L(2,2,2)$ is positive definite on $T_{(2,2)}S$ and we conclude that $(2, 2)$ is a local minimum of f on S.

4. Economic interpretation of the Lagrange multiplier

Let the problem with equality constraints

$$\max\ f(x) \quad \text{s.t.: } g(x) = b. \tag{4.1}$$

Definition 4.1. The value function of problem (4.1) is

$$V(b) = \max\{f(x)\ :\ g(x) = b\}, \quad b \in B \subseteq \mathbb{R}^m,\ \text{an open set.}$$

The optimal policy correspondence of problem (4.1) is

$$X_0(b) = \{x_0 \in S\ :\ f(x_0) \geq f(x)\ \forall x \in S\}.$$

In what follows we will suppose that $X_0(b)$ is a singleton, and thus we refer to the optimal correspondence as the optimal policy function and denote it by $x_0(b)$ ($X_0(b) = \{x_0(b)\}$). Note that $V(b) = f(x_0(b))$ by definition. How does V depends on b?

In the following theorem, recall that the Lagrangian is defined as $L(x, \lambda) = f(x) + \lambda \cdot (b - g(x))$.

Theorem 4.2. Let $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$, such that f and g are of class C^2 in D. Let $b_0 \in B$ such that $x_0(b_0)$ is a regular solution of (1.1) such that

$$\begin{vmatrix} (0)_{m\times m} & \mathrm{D}\, g(x_0(b_0)) \\ \mathrm{D}\, g(x_0(b_0))^t & \mathrm{H}_x\, L(x_0(b_0), \lambda_0) \end{vmatrix} \neq 0,$$

where λ_0 is a vector of Lagrange multipliers associated to $x_0(b_0)$. The, the value function V is of class C^1 in a ball centered at b_0 and

$$\lambda_i^0 = \frac{\partial V}{\partial b_i}(b_0), \quad i \in \{1, \dots, m\}. \tag{4.2}$$

Remark 4.3. The Lagrange multiplier λ_i is associated to the i-th constraint. It measures how sensitive the optimal value is to changes in the independent term b_i. It is for this reason that λ_i is called the **shadow price** or **marginal value** of b_i. Note that an increment of one unit in b_i 'ceteris paribus' has an effect on the optimal value that can be estimated as

$$V(b_1^0, \dots, b_{i-1}^0, b_i^0 + 1, b_{i+1}^0, \dots, b_m^0) - V(b) \approx \lambda_i.$$

In general, if the increment in b_0 is Δb (preferably small), then

$$V(b_0 + \Delta b) - V(b_0) \approx \lambda \cdot \Delta b.$$

Example 4.4 (Indirect utility). Consider the problem,

$$\begin{array}{rl} \max & u(x,y) \\ \text{subject to:} & p_1 x + p_2 y = m \end{array}$$

In this problem a consumer chooses bundles of consumption (x, y) subject to the constraint that, given the prices (p_1, p_2), this bundle costs $p_1 x + p_2 y = m$ and the income of the agent is m.
To solve this problem we consider the Lagrangian function,

$$L(x) = u(x) + \lambda(m - p_1 x - p_2 y)$$

Let $x(p_1, p_2, m)$, $y(p_1, p_2, m)$ be the solution (assume it is unique). Let,

$$V(p_1, p_2, m) = u(x(p_1, p_2, m))$$

be the indirect utility. Then,

$$\frac{\partial V}{\partial m} = \lambda$$

Thus, λ represents the marginal utility of income.

Example 4.5. A consumer solves the problem max xy, subject to $2x + 4y = 32$, $x, y \geq 0$ and find out that $\lambda_0 = 2$. Which is the increment of utility of the consumer if she has income 33 instead of 32?
We know that the marginal value of income for the consumer when the income is 32 is $V'(32) = \lambda_0 = 2$. Hence, we can estimate that the increment in utility is $\Delta V = V(33) - V(32) \approx 2$.

Remark 4.6. Let us justify equation (4.2) for the case in which there is only one restriction.

$$\begin{array}{rl} \max & f(x) \\ \text{s.t.:} & g(x) = b \end{array}$$

Let λ be the Lagrange multiplier and let $x(b)$ be the solution to the above problem. Then, the Lagrange equations are

$$\frac{\partial f}{\partial x_k} = \lambda \frac{\partial g}{\partial x_k} \qquad k = 1, \dots, n.$$

On the one hand, since $x(b)$ satisfies the equation

$$g\left(x(b)\right) = b$$

at $b = b_0$. The condition of the Theorem allows us to apply the Implicit Function Theorem, so $x(b)$ is well defined and of class C^1 in an interval centered at b_0. Differentiating the above equation we obtain

$$\sum_{k=1}^{n} \frac{\partial g}{\partial x_k}\left(x(b)\right) \frac{\partial x_k}{\partial b}(b) = 1.$$

On the other hand, using the chain rule

$$\frac{\partial f\left(x(b)\right)}{\partial b} = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\left(x(b)\right) \frac{\partial x_k}{\partial b}(b) = \lambda \sum_{k=1}^{n} \frac{\partial g}{\partial x_k}\left(x(b)\right) \frac{\partial x_k}{\partial b}(b) = \lambda$$

Remark 4.7. Equation (4.2) also holds for inequality constraints, as we will see.