10.018 Modelling Space and Systems

Cohort 4.1

Lagrange Multipliers

Term 2, 2021

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Before we start....

To get the most out of this cohort, you should already be familiar with

1. Lagrange multipliers in lecture

as we will be going through

1. Using the Lagrange multiplier for optimization

Example

A box is made of cardboard with double-thick sides, a triple-thick bottom, a single-thick front and back and no top. Its volume is fixed at 3. What is the minimum area of cardboard used to construct this box? We want to do three things

(1) We want to formulate the problem of finding the dimensions of the box using the minimum amount of material as a **constrained** optimization problem, i.e.

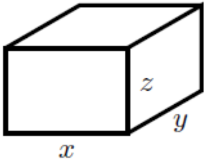
$$\min f(\vec{x}), \quad \text{subject to} \quad \vec{x} \in R \subset \mathbb{R}^n$$

(2) Classify the region R (bounded or unbounded? can we use EVT?)

(3) Solve the optimization problem

Example: Finding the optimization function

The box has dimensions x, y, and z (inedpendent variables).



The function to optimize is the total cardboard used:

$$f(x,y,z) = 2yz + 2yz + 3xy + 2xz = 4yz + 3xy + 2xz.$$

The fixed volume acts as the constraint:

$$g(x,y,z) = xyz - 3 = 0.$$

We obtain the following problem:

$$\left.\begin{array}{lrcl} \min & 4yz + 3xy + 2xz & & \\ \text{s.t.} & xyz & = & 3 \\ & x, y, z & \geq & 0. \end{array}\right\}$$

Example: Identifying the region R

$$\left.\begin{array}{lrcl} \min & 4yz + 3xy + 2xz & & \\ \text{s.t.} & xyz & = & 3 \\ & x, y, z & \geq & 0. \end{array}\right\}$$

The region is closed since it is the intersection of four closed sets: $\{x \geq 0\}, \{y \geq 0\}, \{z \geq 0\}, \{xyz = 3\}$.

It is easy to see that it is unbounded as $(k, \frac{3}{k}, 1)$ satisfies constraints for any $k > 0$.

Since it is unbounded EVT cannot be applied.

Lecture: Necessary Conditions for Optimality

From the Lecture:
Lagrange multipliers for problems with **one constraint**.

Theorem (Necessary condition for optimality)

Let $\vec{x}^*$ be a local maximum (or minimum) of f subject to $g(\vec{x}) = 0$.
Then there exists a unique **Lagrange multiplier** $\lambda^* \in \mathbb{R}$ such that:

$$\nabla f(\vec{x}^*) + \lambda^* \nabla g(\vec{x}^*) = \vec{0}$$

whenever $\nabla g(\vec{x}^*) \neq \vec{0}$ and $\vec{x}^*$ is not a boundary point of the constraint.

Hence each local optimum $\vec{x}^*$ of f subject to $g(\vec{x}) = 0$ satisfies

$$\begin{cases} \nabla f(\vec{x}^*) + \lambda \nabla g(\vec{x}^*) & = & \vec{0} \\ g(\vec{x}^*) & = & 0 \end{cases}$$

whenever $\nabla g(\vec{x}^*) \neq \vec{0}$ and $\vec{x}^*$ is not on the boundary of the constraint.

Example: Solving the constrained optimization problem

The formulation has one equality constraint, and some inequality constraints to indicate that the variables are nonnegative.

Today we discuss a method to deal with equality constraints. What to do with inequalities, in particular the commonly used nonnegativity constraints?

The approach we adopt is to drop nonnegativity conditions. We obtain an equality constrained problem that is called the relaxed problem.

If the optimal solution for the relaxed problem is nonnegative, then it is also optimal for the original problem.

Example: Solving the constrained optimization problem

$$\left.\begin{array}{lrcl} \min & 4yz + 3xy + 2xz & & \\ \text{s.t.} & xyz & = & 3 \\ & x, y, z & \geq & 0. \end{array}\right\}$$

Step 1: Compute the gradients of f and g:

$$\begin{aligned} \nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] &= [3y + 2z, 4z + 3x, 4y + 2x] \\ \nabla g = \left[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}\right] &= [yz, xz, xy]. \end{aligned}$$

Step 2: Setup the system with necessary optimality conditions, $\nabla f + \lambda \nabla g = \vec{0}$ and $g(x, y, z) = 0$:

$$\begin{aligned} 3y + 2z + \lambda yz &= 0 \\ 4z + 3x + \lambda xz &= 0 \\ 4y + 2x + \lambda xy &= 0 \\ xyz &= 3 \end{aligned}$$

Example: Solving the constrained optimization problem

Since $x, y, z > 0$ we obtain:

$$\frac{3}{z} + \frac{2}{y} = \frac{4}{x} + \frac{3}{z} = \frac{4}{x} + \frac{2}{y} = -\lambda.$$

Hence $x = 2y, z = \dfrac{3y}{2}$.

Since $xyz = 3$ we obtain $3y^3 = 3 \implies y = 1$. Putting everything together, we find $\lambda = -4$ and the point $(x, y, z) = (2, 1, 3/2)$ with objective value 18.

This is the only candidate to be a local minimum.

There are various ways to show that this is indeed the global minimum, such as the AM-GM inequality, or arguing from context of problem.

Application - power systems operation

For a power system operator with several generation units, the goal of economic dispatch is to determine an optimal generation schedule that minimizes operating cost. It is a problem at the intersection of engineering and economics.

Main idea: in order to serve a load at minimum total cost, generators with the lowest marginal costs must be used first, subject to operational constraints.

Power systems operation: model

You are a power system operator with three generators. The load to be served is 952 MW. The cost functions of the generators are:

$$f_1(x_1) = x_1 + \frac{1}{16}\, x_1^2$$
$$f_2(x_2) = x_2 + \frac{1}{80}\, x_2^2$$
$$f_3(x_3) = x_3 + \frac{1}{40}\, x_3^2$$

where f_i is the cost per hour (\$/h) and x_i is the output power (MW) of generator i for $i = 1, 2, 3$.

Let us formulate an optimization problem to determine the optimal allocation of generators, i.e. What is the minimum cost per hour to output a total of 952MW?

Activity 1 (15 min)

The optimization problem is:

$$
\begin{array}{lrcl}
\min & f_1(x_1) + f_2(x_2) + f_3(x_3) & = & \\
 & x_1 + \frac{1}{16}x_1^2 + x_2 + \frac{1}{80}x_2^2 + x_3 + \frac{1}{40}x_3^2 & & \\
\text{s.t.} & x_1 + x_2 + x_3 - 952 & = & 0 \\
 & x_1, x_2, x_3 & \geq & 0.
\end{array}
$$

(a) Use the method of Lagrange multipliers to find the optimal solution to this problem.

(b) Find the value of λ. What are the units that lambda is measured in?

Activity 1 (solutions)

We ignore non-negativity constraints $x_1, x_2, x_3 \geq 0$ (**relaxation**).

Lagrange multipliers method: $\nabla f + \lambda \nabla g = \vec{0}$ plus the problem's constraints yield:

$$\begin{aligned} 1 + \tfrac{1}{8}x_1 + \lambda &= 0 \\ 1 + \tfrac{1}{40}x_2 + \lambda &= 0 \\ 1 + \tfrac{1}{20}x_3 + \lambda &= 0 \\ x_1 + x_2 + x_3 - 952 &= 0 \end{aligned}$$

We find $x_1 = 112$ MW, $x_2 = 560$ MW, $x_3 = 280$ MW, $-\lambda = 15$ \$/MWh. The total cost f is 7,616 \$/h. (We can see that λ is expressed in \$/MWh by doing dimensional analysis.)

This is the generation schedule that minimizes the hourly cost of production. (You can show that it is a global minimum but we do not require it here).

Activity 1 (interpretation)

We claim that the negative of the Lagrange multiplier $-\lambda$ is the rate at which the optimal value of the objective function f changes if we change the right-hand side of the constraint.

For this problem, $-\lambda$ is the incremental cost of production, giving the power system operator a cut-off price for potentially buying (or selling) generation. If they can purchase power generation for less than $-\lambda$, it is an opportunity worth pursuing because it is less expensive than their marginal production costs. In fact, $-\lambda$ is expressed in \$/MWh.

We now study why the Lagrange multiplier has this interpretation.

Interpretation of Lagrange Multipliers

Consider the equality constrained problem with a single constraint:

$$\begin{array}{rl} \min & f(\vec{x}) \\ \text{s.t.} & g(\vec{x}) = 0. \end{array}$$

Denote the optimal solution by $\vec{x}^* = (x^*, y^*)$.

To interpret λ, we look at **how the optimum value of the function f changes as the value constraint function g is varied**.

$$g(\vec{x}) = 0 \quad \text{changes to} \quad g(\vec{x}) = c$$

In general, the optimum point (x^*, y^*) depends on the constraint value c. So, provided x^* and y^* are differentiable functions of c, we use the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}c} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x^*}{\mathrm{d}c} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y^*}{\mathrm{d}c}$$

Interpretation of λ

$$\frac{\mathrm{d}f}{\mathrm{d}c} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x^*}{\mathrm{d}c} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y^*}{\mathrm{d}c}$$

At the optimum point (x^*, y^*), we have $f_x = -\lambda g_x$ and $f_y = -\lambda g_y$ (**self-check: why?**), and therefore

$$\frac{\mathrm{d}f}{\mathrm{d}c} = -\lambda\left(\frac{\partial g}{\partial x}\frac{\mathrm{d}x^*}{\mathrm{d}c} + \frac{\partial g}{\partial y}\frac{\mathrm{d}y^*}{\mathrm{d}c}\right) = -\lambda\frac{\mathrm{d}g}{\mathrm{d}c}$$

But, since $g(x^*(c), y^*(c)) = c$, we see that $\mathrm{d}g/\mathrm{d}c = 1$, so we have

$$\frac{\mathrm{d}f}{\mathrm{d}c} = -\lambda.$$

Or if change c in RHS is small:

$$f_{\text{new opt}} - f_{\text{old opt}} \approx (-\lambda)\,\Delta c$$

Interpretation of Lagrange multipliers

Interpretation: up to the first order, the change in objective function value is equal to the negative of the Lagrange multiplier times the change in the right hand side value c.

Remark

Let $f_{\text{opt}}(c)$ be the optimal value of the problem:

$$\min f(\vec{x}), \quad \text{subject to } g(\vec{x}) = c,$$

expressed as a function of c. From the previous slide $-\lambda = \frac{\mathrm{d}f_{\text{opt}}}{\mathrm{d}c}$.

If there are multiple constraints, each Lagrange multiplier indicates the rate of change of the optimal objective function value with respect to the change of the rhs value of the corresponding constraint.

Example: in the power systems operation model, $-\lambda$ indicates the increase in cost per unit of load to be served.

Interpretation of Lagrange multipliers

Remark

Let $f_{\text{opt}}(c)$ be the optimal value of the problem:

$$\min f(\vec{x}), \qquad \text{subject to } g(\vec{x}) = c,$$

expressed as a function of c. From the previous slide $-\lambda = \frac{\mathrm{d} f_{\text{opt}}}{\mathrm{d}c}$.

Example: In Activity 1, the total cost f is 7,616 \$/h, and $-\lambda = 15$ \$/MWh indicates the increase in cost per unit of load to be served.

So if we need to serve 962MWh (instead of 952MWh) the cost will increase **approximately** by $15 \cdot 10 = 150$ \$/h.

Break

5 min break

Don't be late.

Economics application - maximizing production

The Cobb-Douglas production function is a standard model to relate total production (measured in dollars), labor and capital inputs.

Under this model, the production $f(x, y)$ (measured in dollars) resulting from x units of capital and y units of labor is defined as:

$$f(x, y) = A\, x^{\alpha} y^{\beta}, \quad \text{where} \quad \alpha + \beta = 1.$$

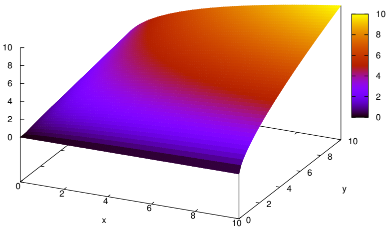
Here A, α and β are positive numbers, and are given parameters of the model.

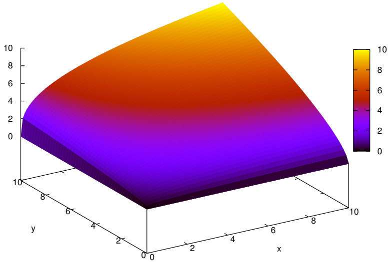
(Alternative models consider $\alpha + \beta \neq 1$, but here we focus on the case $\alpha + \beta = 1$.)

Examples of production function

Under the Cobb-Douglas model, multiplying both capital and labor k times multiplies the production k times.

$$f(kx, ky) = A(kx)^{\alpha}(ky)^{\beta} = Ak^{\alpha+\beta}x^{\alpha}y^{\beta} = Akx^{\alpha}y^{\beta}.$$





This is an example with $A = 1, \alpha = 0.35, \beta = 0.65$.

Activity 2 (15 minutes)

Suppose each unit of capital costs $\$a$ and each unit of labor costs $\$b$. Our total budget is $\$c$. What is the maximum production under a budget of c?

We can write the problem of maximizing production as:

$$\left.\begin{array}{rl} \max f(x,y) = & Ax^\alpha y^\beta \\ \text{s.t. : } g(x,y) = & ax + by - c = 0. \\ & x, y \geq 0. \end{array}\right\}$$

Use the method of Lagrange multipliers to show that the maximum of this problem is achieved when

$$x = \frac{\alpha c}{a} \quad \text{and} \quad y = \frac{\beta c}{b}.$$

Is this a global maximum?

Activity 2 (solutions)

Let us first ignore the non-negativity constraints (relaxation of the original problem). Compute the gradients:

$$\nabla f = \begin{bmatrix} A\alpha x^{\alpha-1}y^{\beta} \\ A\beta x^{\alpha}y^{\beta-1} \end{bmatrix} \qquad \nabla g = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Lagrange multipliers method: $\nabla f + \lambda \nabla g = \vec{0}$ plus the problem's constraints:

$$\begin{aligned} A\alpha x^{\alpha-1}y^{\beta} + \lambda a &= 0 \\ A\beta x^{\alpha}y^{\beta-1} + \lambda b &= 0 \\ ax + by - c &= 0. \end{aligned}$$

From the first two equations we get:

$$\lambda = -\frac{A\alpha x^{\alpha-1}y^{\beta}}{a} = -\frac{A\beta x^{\alpha}y^{\beta-1}}{b}.$$

Activity 2 (solutions)

$$\lambda = -\frac{A\alpha x^{\alpha-1}y^{\beta}}{a} = -\frac{A\beta x^{\alpha}y^{\beta-1}}{b}.$$

Dividing by $Ax^{\alpha}y^{\beta}$, we get:

$$y = \frac{\beta a x}{\alpha b}.$$

Plugging into $ax + by - c = 0$, we obtain:

$$x = \frac{\alpha c}{a(\alpha+\beta)} = \frac{\alpha c}{a} \qquad \text{so that} \qquad y = \frac{\beta c}{b}.$$

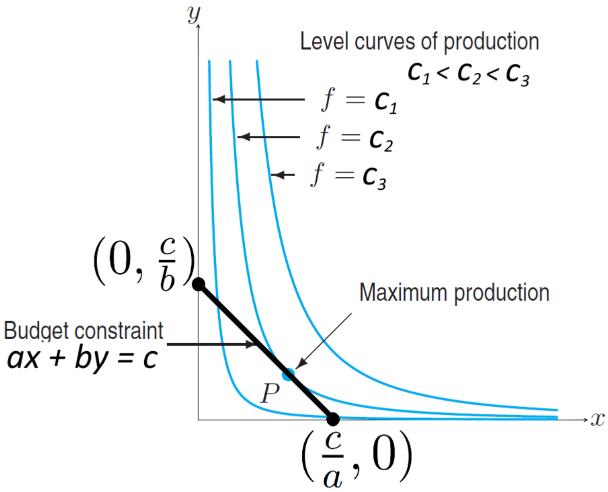
Note that this solution satisfies the original constraints since both x and y are nonnegative.

Activity 2 (solutions)

We can see that this must be the maximum using the EVT since the region of the original problem (before dropping the nonnegativity constraints) is closed and bounded and the objective function is continuous.

The maximum is either the point we have calculated or the two end points of the region. But the end points $(\frac{c}{a}, 0)$ and $(0, \frac{c}{b})$ have objective function value equal to 0 and therefore they are global minima of the function over the region.

Activity 2 (solutions)



The Multinomial Distribution

In applications with more than two outcomes, we may want to estimate the probability that each outcome occurs. Here are two examples.

- In Natural Language Processing (NLP), we might want to look at the distribution of **words** for a specific **topic**. E.g., for the topic **Higher Education**, the word **university** might occur 5% of the time.
- In election polling (with more than two candidates), we might want to look at the chances of a particular candidate winning the election.

Set Up

Suppose we have K outcomes (think of this as K words, or K candidates).

We denote n_i to be the number of times we see the i^{th} outcome, with $n_1 + n_2 + \ldots + n_K = n$.

We want to estimate the **probability** p_i of observing the i^{th} outcome. Note that $0 \le p_i \le 1$ since p_i's are probabilities, and they should add up to 1.

The likelihood function (recall Modelling and Analysis) is given as

$$f(p_1, p_2, \ldots, p_K) = C \prod_{i=1}^{K} p_i^{n_i}, \text{subject to } \sum_{i=1}^{K} p_i = 1$$

which we will simplify for you on the next slide.

Self Check: What happens when $p_i = 0$?

You will learn how to derive this in a probability / statistics class.

Set Up

The general log-likelihood function can be written as

$$f(p_1, p_2, \ldots, p_K) = \sum_{i}^{K} n_i \log(p_i), \text{subject to } \sum_{i=1}^{K} p_i = 1$$

when all $p_i > 0$.

Self Check: Why do we take logs?

Activity 3 (10 min)

The method of maximum likelihood estimation involves **maximizing** the likelihood function given by

$$f(p_1, p_2, \ldots, p_K) = \sum_{i=1}^{K} n_i \log(p_i), \text{subject to } \sum_{i=1}^{K} p_i = 1$$

Find the optimal values of $p_1, \ldots, p_k$ using Lagrange multipliers.

Activity 3 (solutions)

The gradients of the objective function and constraint are:

$$\nabla f = \begin{bmatrix} \frac{n_1}{p_1} \\ \frac{n_2}{p_2} \\ \vdots \\ \frac{n_K}{p_K} \end{bmatrix} \qquad \nabla g = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Activity 3 (solutions)

Set up a system using the Lagrange multipliers method:

$$\begin{aligned} \frac{n_1}{p_1} + \lambda &= 0 \\ &\vdots \\ \frac{n_K}{p_K} + \lambda &= 0 \\ \sum_{i=1}^{K} p_i &= 1 \end{aligned}$$

All K equations yield $p_i = \dfrac{n_i}{n}$.

Interpretation: If we see that the i^{th} outcome has n_i observations out of n total observations, then it is reasonable to estimate that the probability of the i^{th} outcome happening is $\frac{n_i}{n}$.

It can be shown that this is a global maximum, but we will not cover this proof in class.

Summary

We have covered:

- Application of the method of Lagrange multipliers in power systems, economics and probability.
- Examples of how the method can be used to find formulas for the solutions, and analyze them.
- Interpretation of Lagrange multipliers in terms of sensitivity of the optimal solution.

Textbook: read Section 19.8, then try some of Exercises 19.8.1–19.8.23. You may discuss them on Piazza.