MATH 3290 Mathematical Modeling

Chapter 13: Optimization of Continuous Models

Kuang HUANG

April 16, 2024

Department of Mathematics
The Chinese University of Hong Kong

Course webpage

https://www.math.cuhk.edu.hk/course/2324/math3290

SCAN ME

About Final

- Date: Apr. 25.
- The exam is a closed-book 2-hour exam.
- Laptops, tablets, and smartphones are not permitted; however, calculators are allowed.
- Review classes on Apr. 19.

Scope of Final

- Chapter 1: Modeling Change (difference equations)
- Chapter 3: Model Fitting (Chebeshev criterion, least-squares criterion)
- Chapter 4: Experimental Modeling (one-term models, high-order polynomial models, cubic splines)
- Chapter 7: Optimization of Discrete Models (linear programming)
- Chapter 8: Modeling Using Graph Theory (shortest path problem, maximal flow problem)

- Chapter 11: Modeling with a Differential Equation (solving the equation, equilibrium points and their stability, graphical method, Euler's method)
- Chapter 12: Modeling with Systems of Differential Equations (solving the system of equations, equilibrium points and their stability, graphical method, Euler's method)
- Chapter 13: Optimization of Continuous Models (nonlinear optimization, unconstrained optimization, equality/inequality constraints, KKT condition)

Introduction

We will consider optimization problems in which the objective function f is nonlinear.

That is, find X^* such that

$$f(X) \quad \text{is optimized.}$$

Again, $X = (X_1, \ldots, X_n)$ are called decision variables.

- **Unconstrained:** f is optimized without restrictions on X ($X \in \mathbb{R}^n$).
- **Constrained:** there are restrictions on X.
  - **Equality**, $g_i(X) = b_i$, for $i = 1, 2, \ldots, m$.
  - **Inequality**, $g_i(X) \leq b_i$, for $i = 1, 2, \ldots, m$.
  - **Mixed**, both equality and inequality.

Unconstrained optimization

We find (x^*, y^*) such that $f(x, y)$ is optimized.

Method 1: using critical points.

If (x^*, y^*) attains the maximum/minimum of $f(x, y)$, then

$$\nabla f(x^*, y^*) := \big(f_x(x^*, y^*), f_y(x^*, y^*)\big) = (0, 0)$$

To check (x^*, y^*) is a max or min, we use the second derivative test.

We define the Hessian matrix by

$$H(x, y) = \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{pmatrix}.$$

Taylor expansion:

$$
\begin{aligned}
f(x^* + tu_1, y^* + tu_2) = & f(x^*, y^*) + (tu_1) f_x(x^*, y^*) + (tu_2) f_y(x^*, y^*) \\
& + \frac{1}{2}(tu_1)^2 f_{xx}(x^*, y^*) + (tu_1)(tu_2) f_{xy}(x^*, y^*) \\
& + \frac{1}{2}(tu_2)^2 f_{yy}(x^*, y^*) + O(t^3).
\end{aligned}
$$

Define $u = (u_1, u_2)$. Note that $\nabla f(x^*, y^*) = (0, 0)$.

The above formula can be written as

$$
f(x^* + tu_1, y^* + tu_2) = f(x^*, y^*) + \frac{t^2}{2} u^T H(x^*, y^*) u + O(t^3).
$$

$$f(x^* + tu_1, y^* + tu_2) = f(x^*, y^*) + \frac{t^2}{2} u^T H(x^*, y^*) u + O(t^3)$$

- If $H(x^*, y^*)$ is positive definite, that is,

$$u^T H(x^*, y^*) u > 0, \qquad \text{for all non-zero } u.$$

(also equivalent to $H(x^*, y^*)$ has positive eigenvalues), then

$$f(x^* + tu_1, y^* + tu_2) \geq f(x^*, y^*), \quad \text{for small } t > 0.$$

Hence, (x^*, y^*) is a local min.

- If $H(x^*, y^*)$ is negative definite, that is,

$$u^T H(x^*, y^*) u < 0, \qquad \text{for all non-zero } u.$$

(also equivalent to $H(x^*, y^*)$ has negative eigenvalues), then

$$f(x^* + tu_1, y^* + tu_2) \leq f(x^*, y^*), \quad \text{for small } t > 0.$$

Hence, (x^*, y^*) is a local max.

Example: maximizing profit

Assume you are producing computers.

1. Two specs: one with 27 inch monitor, the other with 31 inch monitor.
2. A fixed cost: 400,000.
3. The cost for making one 27 (31) inch model is 1950 (2250).
4. The retail price for 27 (31) model is 3390 (3990).
5. For each unit sold, the price is reduced by 0.1.
6. For each 27 model sold, the price of 31 model is reduced by 0.04.
7. For each 31 model sold, the price of 27 model is reduced by 0.03.

We can then set up the following notations.

- $x_1, x_2 =$ numbers of 27 (31) inch models.
- $P_1, P_2 =$ prices of 27 (31) inch models.

$$P_1 = 3390 - 0.1x_1 - 0.03x_2, \quad P_2 = 3990 - 0.04x_1 - 0.1x_2.$$

- $R =$ revenue obtained from sales $= P_1x_1 + P_2x_2$.
- $C =$ cost to make computers $= 400{,}000 + 1950x_1 + 2250x_2$.
- $P =$ total profit $= R - C$.

Let us forget about non-negativity constraints.

Combining above, we will maximize

$$
\begin{aligned}
P(x_1, x_2) &= R - C = P_1 x_1 + P_2 x_2 - C \\
&= 1440 x_1 - 0.1 x_1^2 + 1740 x_2 - 0.1 x_2^2 - 0.07 x_1 x_2 - 400{,}000.
\end{aligned}
$$

Finding partial derivatives,

$$
\frac{\partial P}{\partial x_1} = 1440 - 0.2 x_1 - 0.07 x_2, \quad \frac{\partial P}{\partial x_2} = 1740 - 0.07 x_1 - 0.2 x_2.
$$

Setting partial derivatives to zero, we get $x_1 = 4736$ and $x_2 = 7043$.

To find the Hessian matrix, we compute second derivatives

$$
\frac{\partial^2 P}{\partial x_1^2} = -0.2, \quad \frac{\partial^2 P}{\partial x_1 x_2} = -0.07, \quad \frac{\partial^2 P}{\partial x_2^2} = -0.2.
$$

Hence, the Hessian matrix is

$$H = \begin{pmatrix} -0.2 & -0.07 \\ -0.07 & -0.2 \end{pmatrix}.$$

Note, it is independent of (x, y) for this example.

To find eigenvalues, we set $\det(H - \mu I) = 0$, which implies

$$(0.2 + \mu)^2 - 0.07^2 = 0.$$

So,

$$\mu = -0.2 \pm 0.07.$$

Hence, all eigenvalues of H are negative.

We conclude that the point $(4736, 7043)$ is a max.

The gradient method

Note that, to find critical points, we need to solve a nonlinear system

$$f_x(x^*, y^*) = 0, \qquad f_y(x^*, y^*) = 0.$$

This may not be easy.

Method 2: the gradient method.

To motivate the idea, we recall the definition of directional derivatives.

Let $u = (u_1, u_2)$ be a unit vector. The derivative in the direction u is

$$\frac{\partial f}{\partial u}(x, y) = \lim_{h \to 0^+} \frac{f(x + hu_1, y + hu_2) - f(x, y)}{h}.$$

Note that it is the rate of change of f in the direction u.

From elementary calculus,

$$\frac{\partial f}{\partial u}(x,y) = \nabla f(x,y) \cdot u = |\nabla f(x,y)|\,|u| \cos(\theta),$$

where θ is the angle between $\nabla f(x,y)$ and u.

Since u is a unit vector,

$$\frac{\partial f}{\partial u}(x,y) = |\nabla f(x,y)| \cos(\theta).$$

- The change is largest when $\theta = 0$, that is, when u has the same direction as $\nabla f(x,y)$.
- The change is smallest (most negative) when $\theta = \pi$, that is, when u has the opposite direction as $\nabla f(x,y)$.

The above observations suggest the following method.

Step 1: initialize, choose an initial point (x_0, y_0).

Step 2: move to a better point.

Assume that the current point is (x_k, y_k). How to find a point (x_{k+1}, y_{k+1}) that gives a better value of f ?

- To find the max of f, we should move in the direction $\nabla f(x_k, y_k)$.
- To find the min of f, we should move in the direction $-\nabla f(x_k, y_k)$.

- To find the max of f, we should move in the direction $\nabla f(x_k, y_k)$:

$$x_{k+1} = x_k + \lambda_k f_x(x_k, y_k),$$
$$y_{k+1} = y_k + \lambda_k f_y(x_k, y_k),$$

  where $\lambda_k > 0$ is the distance traveled in the direction $\nabla f(x_k, y_k)$.

- To find the min of f, we should move in the direction $-\nabla f(x_k, y_k)$:

$$x_{k+1} = x_k - \lambda_k f_x(x_k, y_k),$$
$$y_{k+1} = y_k - \lambda_k f_y(x_k, y_k),$$

  where $\lambda_k > 0$ is the distance traveled.

Step 3: repeat until $\nabla f(x_k, y_k)$ is small.

We still need to determine λ_k (called step size, learning rate in Machine Learning).

Common options:

- Take λ_k as a constant (need to be carefully chosen).
- Using an optimal choice of λ_k (not always available).

For example, to find max value of f, we have

$$x_{k+1} = x_k + \lambda_k f_x(x_k, y_k),$$
$$y_{k+1} = y_k + \lambda_k f_y(x_k, y_k).$$

We then take λ_k such that $f(x_{k+1}, y_{k+1})$ is maximized.

We maximize $g(\lambda) = f(x_k + \lambda f_x(x_k, y_k), y_k + \lambda f_y(x_k, y_k))$.

Example: minimize $f(x, y) = x^3 - 2x + y^2$.

First, we have $\nabla f(x, y) = (3x^2 - 2, 2y)$. The gradient method is

$$x_{k+1} = x_k - \lambda_k f_x(x_k, y_k) = x_k - \lambda_k(3x_k^2 - 2),$$
$$y_{k+1} = y_k - \lambda_k f_y(x_k, y_k) = y_k - \lambda_k(2y_k).$$

Let

$$g(\lambda) = (x_k - \lambda(3x_k^2 - 2))^3 - 2(x_k - \lambda(3x_k^2 - 2)) - (y_k - \lambda(2y_k))^2.$$

Then we have

$$g'(\lambda) = 3(x_k - \lambda(3x_k^2 - 2))^2(2 - 3x_k^2) + 2(3x_k^2 - 2) + 2(y_k - \lambda(2y_k))(2y_k).$$

To find the min of $g(\lambda)$, we need to solve $g'(\lambda) = 0$.

Recall

$$g'(\lambda) = 3(x_k - \lambda(3x_k^2 - 2))^2(2 - 3x_k^2) + 2(3x_k^2 - 2) + 2(y_k - \lambda(2y_k))(2y_k).$$

Suppose that the initial guess is $(x_0, y_0) = (0, 0)$.

To find λ_0, we set $g'(\lambda) = 0$ using the initial conditions, giving

$$24\lambda^2 - 4 = 0, \quad \text{which implies} \quad \lambda = \frac{1}{\sqrt{6}}.$$

Thus, we have $\lambda_0 = 1/\sqrt{6}$. Hence,

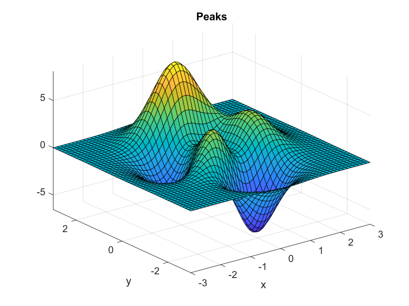
$$x_1 = x_0 - \lambda_0 f_x(x_0, y_0) = -\frac{2}{\sqrt{6}},$$
$$y_1 = y_0 - \lambda_0 f_y(x_0, y_0) = 0.$$

Recall

$$g'(\lambda) = 3(x_k - \lambda(3x_k^2 - 2))^2(2 - 3x_k^2) + 2(3x_k^2 - 2) + 2(y_k - \lambda(2y_k))(2y_k).$$

Next, to see if we need to continue the iteration, we check if $\nabla f(x_1, y_1)$ is small enough.

Note that $\nabla f(x_1, y_1) = (0, 0)$. So we reach the minimum.

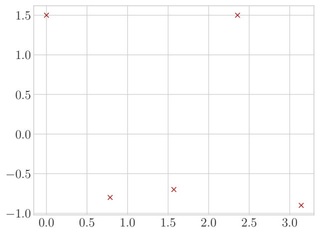


Type in “figure; peaks;” in Matlab

- There are multiple local minima.
- The result of gradient method depends on the initial guess.
- A convex optimization problem is preferable, where local minimum = global minimum.

Example: nonlinear least squares fit

x	0	$\pi/4$	$\pi/2$	$3\pi/4$	π
y	1.5	-0.8	-0.7	1.5	-0.9



The data set shows a periodic trend, it is reasonable to fit a periodic function. Suppose the model function is

$$y = a\cos(bx).$$

We determine the parameters a and b by the least-squares criterion.

$$S(a,b)=\sum_{i=1}^{5}\Big(y_i-a\cos(bx_i)\Big)^2.$$

We will minimize S by the gradient method.

Let a_0 and b_0 are given. Then

$$a_{k+1}=a_k-\lambda_k\frac{\partial S}{\partial a}(a_k,b_k),$$
$$b_{k+1}=b_k-\lambda_k\frac{\partial S}{\partial b}(a_k,b_k).$$

where

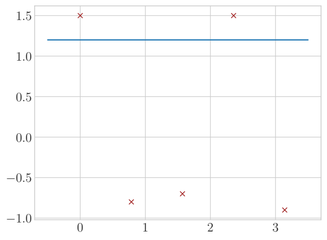
$$\frac{\partial S}{\partial a}=\sum_{i=1}^{5}-2\cos(bx_i)\Big(y_i-a\cos(bx_i)\Big),$$
$$\frac{\partial S}{\partial b}=\sum_{i=1}^{5}2ax_i\sin(bx_i)\Big(y_i-a\cos(bx_i)\Big).$$

We take $a_0 = b_0 = 0.0$ and $\lambda_k = 0.01$.

Iteration	a	b	$\nabla S(a,b)$	$S(a,b)$
0	0.000	0.000	$(-1.200, 0.000)$	6.440
1	0.012	0.000	$(-1.080, 0.000)$	6.426
2	0.023	0.000	$(-0.972, 0.000)$	6.415
3	0.033	0.000	$(-0.875, 0.000)$	6.406
4	0.041	0.000	$(-0.787, 0.000)$	6.399
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
148	0.120	0.000	$(-0.000, 0.000)$	6.368
149	0.120	0.000	$(-0.000, 0.000)$	6.368

Hence, we have $a = 0.120$ and $b = 0.0$.

Recall, we have $a = 0.120$ and $b = 0.0$.



$$y = a\cos(bx) = 0.12$$

- The result is not good, it is trapped by a local minimum.
- The function value does not decrease much.

We still take $\lambda_k = 0.01$. Looking at the data, it is reasonable to take

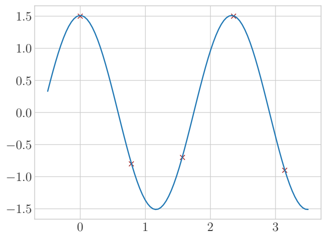
$$a_0 = 1.5, \quad b_0 = 3.$$

(The data set shows almost an amplitude of 1.5 and a period of 3.)

Iteration	a	b	$\nabla S(a,b)$	$S(a,b)$
0	1.500	3.000	$(0.947, 5.929)$	1.111
1	1.491	2.941	$(0.767, 5.258)$	0.772
2	1.483	2.888	$(0.560, 4.620)$	0.506
3	1.477	2.842	$(0.353, 3.886)$	0.306
4	1.474	2.803	$(0.173, 3.075)$	0.169
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
298	1.511	2.701	$(-0.000, 0.000)$	0.001
299	1.511	2.701	$(-0.000, 0.000)$	0.001

Hence, we have $a = 1.511$ and $b = 2.701$.

Recall, we have $a = 1.511$ and $b = 2.701$.



$$y = a\cos(bx) = 1.511\cos(2.701x)$$

- The result is very good, global minimum reached.
- The choice of initial guess is very important.

Constrained optimization

We first consider problems with equality constraints.

We find X^* in an open set S that

$$\begin{aligned} &\text{optimize} && f(X) \\ &\text{subject to} && g(X) = 0. \end{aligned}$$

Roughly speaking, an open set is a region without boundary.

The region $\{x > 0, y > 0\}$ is open.

⚠️ ⚠️ ⚠️ We must priorly know the min is not on the boundary of S.

Method of Lagrange multiplier

Consider

$$\begin{aligned} &\text{optimize} && f(x,y), \\ &\text{subject to} && g(x,y) = 0. \end{aligned}$$

Define

$$L(x,y,\lambda) = f(x,y) + \lambda g(x,y).$$

If (x^*, y^*) optimizes (max or min) f and $\nabla g(x^*, y^*) \neq 0$, then there is λ^* such that

$$L_x(x^*, y^*, \lambda^*) = 0, \; L_y(x^*, y^*, \lambda^*) = 0, \; L_\lambda(x^*, y^*, \lambda^*) = 0.$$

Equivalently,

$$\begin{aligned} &f_x(x^*, y^*) + \lambda^* g_x(x^*, y^*) = 0, \\ &f_y(x^*, y^*) + \lambda^* g_y(x^*, y^*) = 0, \\ &g(x^*, y^*) = 0. \end{aligned}$$

We can solve the above equations for x^*, y^*, and λ^*.

Consider a point $(x, y) = (x^* + \varepsilon, y^* + \eta)$ near (x^*, y^*). We will then check if this point (x, y) will increase or decrease the value of f.

Note that the point (x, y) should satisfy the constraint $g(x, y) = 0$. So,

$$g(x^* + \varepsilon, y^* + \eta) = 0.$$

Thus, η is a function of ε (locally, by the implicit function theorem). Write $\eta = \eta(\varepsilon)$. Note $\eta(0) = 0$.

Let $F(\varepsilon) = f(x^* + \varepsilon, y^* + \eta(\varepsilon))$. Then F has a max or min at $\varepsilon = 0$.

- $F'(0) = 0$ gives the first two equations.
- It is a max if $F''(0) < 0$, and it is a min if $F''(0) > 0$, which determine (x^*, y^*) is a max or min.

How to calculate $F'(0)$ and $F''(0)$?

Take derivative for $F(\epsilon)$ and let $\epsilon = 0$,

$$F'(\epsilon) = f_x(x^* + \epsilon, y^* + \eta(\epsilon)) + f_y(x^* + \epsilon, y^* + \eta(\epsilon))\eta'(\epsilon)$$
$$\Rightarrow F'(0) = f_x(x^*, y^*) + f_y(x^*, y^*)\eta'(0).$$

Take derivatives for $g(x^* + \epsilon, y^* + \eta(\epsilon)) = 0$ on both sides and let $\epsilon = 0$,

$$g_x(x^* + \epsilon, y^* + \eta(\epsilon)) + g_y(x^* + \epsilon, y^* + \eta(\epsilon))\eta'(\epsilon) = 0$$
$$\Rightarrow g_x(x^*, y^*) + g_y(x^*, y^*)\eta'(0) = 0,$$

from which we can solve $\eta'(0)$ by $g_x(x^*, y^*)$ and $g_y(x^*, y^*)$.

Similarly,

$$F''(0) = f_{xx}(x^*, y^*) + 2f_{xy}(x^*, y^*)\eta'(0) + f_{yy}(x^*, y^*)(\eta'(0))^2 + f_y(x^*, y^*)\eta''(0),$$

while $\eta''(0)$ can be solved from

$$0 = g_{xx}(x^*, y^*) + 2g_{xy}(x^*, y^*)\eta'(0) + g_{yy}(x^*, y^*)(\eta'(0))^2 + g_y(x^*, y^*)\eta''(0).$$

Generalization

Consider

$$\begin{aligned} &\text{optimize} && f(x,y,z), \\ &\text{subject to} && g(x,y,z) = 0, \quad h(x,y,z) = 0. \end{aligned}$$

Define

$$L(x,y,z,\lambda,\mu) = f(x,y,z) + \lambda g(x,y,z) + \mu h(x,y,z).$$

If (x^*, y^*, z^*) optimizes f, then there are λ^*, μ^* such that

$$L_x(x^*, y^*, z^*, \lambda^*, \mu^*) = L_y(x^*, y^*, z^*, \lambda^*, \mu^*) = L_z(x^*, y^*, z^* \lambda^*, \mu^*) = 0,$$

$$L_\lambda(x^*, y^*, z^*, \lambda^*, \mu^*) = L_\mu(x^*, y^*, z^*, \lambda^*, \mu^*) = 0.$$

Generalization

Consider

$$\begin{aligned} &\text{optimize} && f(x,y,z), \\ &\text{subject to} && g(x,y,z)=0, \quad h(x,y,z)=0. \end{aligned}$$

Define

$$L(x,y,z,\lambda,\mu) = f(x,y,z) + \lambda g(x,y,z) + \mu h(x,y,z).$$

Equivalently,

$$\begin{aligned} &f_x(x^*,y^*,z^*) + \lambda^* g_x(x^*,y^*,z^*) + \mu^* h_x(x^*,y^*,z^*) = 0, \\ &f_y(x^*,y^*,z^*) + \lambda^* g_y(x^*,y^*,z^*) + \mu^* h_y(x^*,y^*,z^*) = 0, \\ &f_z(x^*,y^*,z^*) + \lambda^* g_z(x^*,y^*,z^*) + \mu^* h_z(x^*,y^*,z^*) = 0, \\ &g(x^*,y^*,z^*) = h(x^*,y^*,z^*) = 0. \end{aligned}$$

An example

Optimize $f(x,y,z) = x^2 + 2y - 2z^2$ subject to

$$g(x,y,z) = 2x - y = 0, \quad h(x,y,z) = x + z - 6 = 0.$$

Define

$$L(x,y,z,\lambda,\mu) = x^2 + 2y - 2z^2 + \lambda(2x - y) + \mu(x + z - 6).$$

Taking derivatives:

$$L_x = 2x + 2\lambda + \mu = 0,\ L_y = 2 - \lambda = 0,\ L_z = -4z + \mu = 0,$$

$$L_\lambda = 2x - y = 0,\ L_\mu = x + z - 6 = 0.$$

Solving it, we get

$$x = 14, y = 28, z = -8,$$

$$\lambda = 2,\ \mu = -32.$$

Next, we check if this is a max or min.

Let $(x^*, y^*, z^*) = (14, 28, -8)$ and

$$x = x^* + \varepsilon,\ y = y^* + \eta,\ z = z^* + \delta.$$

Since (x, y, z) satisfies the constraints,

$$2(x^* + \varepsilon) - (y^* + \eta) = 0,\ (x^* + \varepsilon) + (z^* + \delta) - 6 = 0.$$

We get

$$\eta = 2\varepsilon, \quad \delta = -\varepsilon.$$

Let

$$F(\varepsilon) = f(x, y, z) = (x^* + \varepsilon)^2 + 2(y^* + 2\varepsilon) - 2(z^* - \varepsilon)^2.$$

Then

$$F'(\varepsilon) = 2(x^* + \varepsilon) + 4 + 4(z^* - \varepsilon),$$

$$F''(\varepsilon) = -2.$$

Hence, $F''(0) < 0$. So we have a max.

Example from consumer theory

We consider a utility optimization problem.

- A consumer buys two goods, amount of commodity i is x_i.
- The utility is defined as $u(x_1, x_2) = x_1 x_2$.
- The price of commodity i is $p_i > 0$.
- The consumer has income I.

We have the maximization problem

$$\begin{aligned} \max \quad & u(x_1, x_2) = x_1 x_2, \\ \text{subject to} \quad & p_1 x_1 + p_2 x_2 \leq I, \quad x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

Note that, we have three inequality constraints.

Original problem:

$$\begin{aligned} \max \quad & u(x_1, x_2) = x_1 x_2, \\ \text{subject to} \quad & p_1 x_1 + p_2 x_2 \le I, \quad x_1 \ge 0, x_2 \ge 0. \end{aligned}$$

Sometimes, we can simplify the problem as follows. Note that

- If either x_1 or x_2 is zero, then $u = 0$;
- There exists a point in the feasible region with $u > 0$;
- The optimal solution does not satisfy $p_1 x_1 + p_2 x_2 < I$.

The above problem can then be formulated as

$$\begin{aligned} \max \quad & u(x_1, x_2) = x_1 x_2, \\ \text{subject to} \quad & p_1 x_1 + p_2 x_2 = I, \end{aligned}$$

where the max is found in the open set $\{x_1 > 0, x_2 > 0\}$.

New problem:

$$\begin{aligned} \max \quad & u(x_1, x_2) = x_1 x_2, \\ \text{subject to} \quad & p_1 x_1 + p_2 x_2 = I, \end{aligned}$$

where the max is found in the open set $\{x_1 > 0, x_2 > 0\}$.

Now we can use the Lagrange multiplier method . Let

$$L(x_1, x_2, \lambda) = x_1 x_2 + \lambda(I - p_1 x_1 - p_2 x_2).$$

The solution is

$$x_1 = \frac{I}{2p_1}, \quad x_2 = \frac{I}{2p_2}, \quad \lambda = \frac{I}{2p_1 p_2}.$$

An important remark: using the Lagrange multiplier method by removing non-negativity constraints does not always work.

Consider the following example.

$$\begin{aligned} &\max && u(x_1, x_2) = x_1 + x_2, \\ &\text{subject to} && p_1 x_1 + p_2 x_2 = I, \quad x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

If you remove the non-negativity constraints and use the method of Lagrange multiplier, you obtain

$$\begin{aligned} 1 - \lambda p_1 &= 0, \\ 1 - \lambda p_2 &= 0, \\ p_1 x_1 + p_2 x_2 &= I. \end{aligned}$$

This is an inconsistent system.

Thus, one needs to work with inequality constraints.

Inequality constraints

We consider optimization problems with inequality constraints.

Find (x^*, y^*) in some open set S such that

$$\begin{aligned} \text{maximize} \quad & f(x, y), \\ \text{subject to} \quad & g(x, y) \geq 0. \end{aligned}$$

Define

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$

If (x^*, y^*) is the optimal solution (max), there is λ^* such that

$$L_x(x^*, y^*, \lambda^*) = L_y(x^*, y^*, \lambda^*) = 0$$

and

$$L_\lambda(x^*, y^*, \lambda^*) \geq 0, \quad \lambda^* \geq 0, \quad \lambda^* L_\lambda(x^*, y^*, \lambda^*) = 0.$$

This is called the Karush-Kuhn-Tucker multiplier method. The above relations are called KKT conditions.

An example

$$\begin{aligned} &\text{maximize} && f(x,y) = x^2 - y, \\ &\text{subject to} && g(x,y) = 1 - x^2 - y^2 \geq 0. \end{aligned}$$

Define $L(x, y, \lambda) = x^2 - y + \lambda(1 - x^2 - y^2)$.

The conditions are

$$L_x = 2x - 2\lambda x = 0, \quad L_y = -1 - 2\lambda y = 0,$$

$$L_\lambda = 1 - x^2 - y^2 \geq 0, \quad \lambda \geq 0, \quad \lambda(1 - x^2 - y^2) = 0.$$

From the first equation,

$$2x - 2\lambda x = 0 \quad \rightarrow \quad x = 0 \text{ or } \lambda = 1.$$

$$L_x = 2x - 2\lambda x = 0, \quad L_y = -1 - 2\lambda y = 0,$$
$$L_\lambda = 1 - x^2 - y^2 \geq 0, \quad \lambda \geq 0, \quad \lambda(1 - x^2 - y^2) = 0.$$

- If $\lambda = 1$, then $-1 - 2\lambda y = 0$ implies $y = -1/2$.
  Since $\lambda = 1$, the condition $\lambda(1 - x^2 - y^2) = 0$ implies that

  $$1 - x^2 - y^2 = 0$$

  giving $x = \pm\sqrt{3}/2$. The other two conditions are satisfied.
  Two solutions: $(\sqrt{3}/2, -1/2, 1)$ and $(-\sqrt{3}/2, -1/2, 1)$.
- If $x = 0$, then we must have $\lambda > 0$ otherwise $-1 - 2\lambda y = 0$ is a contradiction. Then $\lambda(1 - x^2 - y^2) = 0$ implies that

  $$1 - x^2 - y^2 = 0$$

  giving $y = \pm 1$. Also, we have $\lambda = -1/(2y)$.
  One solution: $(0, -1, 1/2)$.

Recall

$$\begin{aligned} &\text{maximize} && f(x,y) = x^2 - y, \\ &\text{subject to} && g(x,y) = 1 - x^2 - y^2 \geq 0. \end{aligned}$$

Finally, comparing: $f(\pm\sqrt{3}/2, -1/2) = 5/4$ and $f(0, -1) = 1$.

We see that the points $(\pm\sqrt{3}/2, -1/2)$ attains the max.

Explanation of KKT condition

Recall

$$\begin{aligned} \text{maximize} \qquad & f(x, y), \\ \text{subject to} \qquad & g(x, y) \geq 0. \end{aligned}$$

Define

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$

If (x^*, y^*) is the optimal solution, there is λ^* such that

$$L_x(x^*, y^*, \lambda^*) = L_y(x^*, y^*, \lambda^*) = 0$$

and

$$L_\lambda(x^*, y^*, \lambda^*) \geq 0, \quad \lambda^* \geq 0, \quad \lambda^* L_\lambda(x^*, y^*, \lambda^*) = 0.$$

Note that, either $g(x^*, y^*) > 0$ or $g(x^*, y^*) = 0$.

Case 1: $g(x^*, y^*) > 0$.

The problem can be formulated as: find (x^*, y^*) in the open set defined by $S \cap \{g(x, y) > 0\}$ that

$$\text{maximize } f(x, y).$$

Then we have $\nabla f(x^*, y^*) = 0$. The choice of $\lambda^* = 0$ works.

Case 2: $g(x^*, y^*) = 0$.

The problem can be formulated as: find (x^*, y^*) in the open set S that

$$\begin{aligned} &\text{maximize} && f(x, y), \\ &\text{subject to} && g(x, y) = 0. \end{aligned}$$

The Lagrange multiplier method implies there is a λ^* such that

$$L_x(x^*, y^*, \lambda^*) = L_y(x^*, y^*, \lambda^*) = L_\lambda(x^*, y^*, \lambda^*) = 0.$$

Four of the five KKT conditions are satisfied.

We only need to see why $\lambda^* \geq 0$.

We will show that

$$\nabla f(x^*, y^*) \cdot \nabla g(x^*, y^*) \leq 0.$$

From $L_x(x^*, y^*, \lambda^*) = L_y(x^*, y^*, \lambda^*) = 0$, we have

$$\nabla f(x^*, y^*) + \lambda^* \nabla g(x^*, y^*) = 0$$

Thus,

$$\nabla f(x^*, y^*) \cdot \nabla g(x^*, y^*) + \lambda^* |\nabla g(x^*, y^*)|^2 = 0$$

This implies $\lambda^* \geq 0$. (We assume $\nabla g(x^*, y^*) \neq (0, 0)$, otherwise (x^*, y^*) is a critical point of $g(x, y)$, tricky then.)

To show

$$\nabla f(x^*, y^*) \cdot \nabla g(x^*, y^*) \le 0.$$

Consider the directional derivative in the direction $u = (u_1, u_2)^T$:

$$\frac{\partial f}{\partial u} = \lim_{\epsilon \to 0^+} \frac{f(x^* + \epsilon u_1, y^* + \epsilon u_2) - f(x^*, y^*)}{\epsilon} = \nabla f(x^*, y^*) \cdot u \le 0$$

provided the point $(x^* + \epsilon u_1, y^* + \epsilon u_2)$ lies in the feasible region. i.e., $g(x^* + \epsilon u_1, y^* + \epsilon u_2) \ge 0$.

We can take

$$u = \nabla g(x^*, y^*),$$
$$g(x^* + \epsilon u_1, y^* + \epsilon u_2) \approx g(x^*, y^*) + \epsilon \nabla g(x^*, y^*) \cdot u = \epsilon |\nabla g(x^*, y^*)|^2,$$

since it is pointing into the region $\{x \,|\, g(x) \ge 0\}$.

Mixed constraints

Consider problems with both equality and inequality constraints.

Find (x^*, y^*) in some open set S such that

$$\begin{array}{ll} \text{maximize} & f(x,y) \\ \text{subject to} & g(x,y) = 0, \text{ and } h(x,y) \geq 0. \end{array}$$

Define $L(x,y,\lambda,\mu) = f(x,y) + \lambda g(x,y) + \mu h(x,y)$.

If (x^*, y^*) is the optimal solution, there is λ^*, μ^* such that

$$L_x(x^*, y^*, \lambda^*, \mu^*) = L_y(x^*, y^*, \lambda^*, \mu^*) = 0.$$

For equality constraint: $L_\lambda(x^*, y^*, \lambda^*, \mu^*) = 0$.

For inequality constraint: we have

$$L_\mu(x^*, y^*, \lambda^*, \mu^*) \geq 0, \quad \mu^* \geq 0, \quad \mu^* L_\mu(x^*, y^*, \lambda^*, \mu^*) = 0.$$

An example

Consider a problem with both equality and inequality constraints.

$$\begin{aligned} &\text{maximize} && f(x,y) = xy, \\ &\text{subject to} && x + 2y - 4 = 0 \quad \text{and} \quad x - 3 \geq 0. \end{aligned}$$

Define $L(x, y, \lambda, \mu) = xy + \lambda(x + 2y - 4) + \mu(x - 3)$.

An example

KKT conditions:

$$L_x = y + \lambda + \mu = 0,\ L_y = x + 2\lambda = 0,\ L_\lambda = x + 2y - 4 = 0,$$

$$L_\mu = x - 3 \geq 0,\ \mu \geq 0,\ \mu(x-3) = 0.$$

Note:

- $\lambda \neq 0$ (otherwise $x = 0$, which contradicts $x \geq 3$).
- $\mu \neq 0$ (otherwise $\mu = 0$, solving first 3 equations yields $x = 2$, which contradicts $x \geq 3$).
- $\mu(x-3) = 0$ implies that $x = 3$.
- $\lambda = -3/2$, $y = 1/2$ and $\mu = 1$.

Optimal solution $(x^*, y^*) = (3, 1/2)$.

Example: portfolio optimization

Suppose that there are n assets. You want to invest a fixed amount of money. How do you allocate your investments?

Let x_i be the portion of money invested in asset i.

Two important factors: return and risk.

- Assume μ_i are the average return of asset i. On average, you have the following return

$$\mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_n x_n.$$

- Risk is typically modeled by a $n \times n$ positive definite matrix Q. The risk is

$$\frac{1}{2} x^T Q x,$$

  where $x = (x_1, x_2, \ldots, x_n)^T$. Risk is large if this number is big.

Two common ways

- We find x_i so that

$$\max \quad \mu_1 x_1 + \cdots + \mu_n x_n - \frac{1}{2} x^T Q x.$$

(maximize return at the same time minimize risk, put weights before return or risk if needed.)
subject to

$$x_1 + \cdots + x_n = 1, \quad x_i \geq 0.$$

- Given a fixed number R, we find x_i

$$\max -\frac{1}{2} x^T Q x$$

subject to

$$x_1 + \cdots + x_n = 1, \quad x_i \geq 0$$

and

$$\mu_1 x_1 + \cdots + \mu_n x_n \geq R.$$

(minimize risk, and having return of at least R.)

Example: consider three assets, that is stocks (S), bonds (B) and money market (M).

Assume the average returns are 10%, 8% and 6%. The risk Q is

	S	B	M
S	2	0.5	0.01
B	0.5	1	−0.01
M	0.01	−0.01	0.1

We need a return of at least 7%. How do you allocate your money?

Let x_1, x_2 and x_3 be the portion of money invested in stock, bond, and money market respectively.

The above problem can be formulated as

$$\max \quad -\frac{1}{2}x^T Qx \left\{ = -\frac{1}{2}(2x_1^2 + x_2^2 + 0.1x_3^2 + x_1x_2 - 0.02x_2x_3 + 0.02x_1x_3) \right\}$$

subject to

$$x_1 + x_2 + x_3 = 1, \quad 10x_1 + 8x_2 + 6x_3 \geq 7, \quad x_i \geq 0.$$

Let

$$\begin{aligned} & L(x_1, x_2, x_3, \lambda, \mu_0, \mu_1, \mu_2, \mu_3) \\ = & -\frac{1}{2}(2x_1^2 + x_2^2 + 0.1x_3^2 + x_1x_2 - 0.02x_2x_3 + 0.02x_1x_3) \\ & + \lambda(x_1 + x_2 + x_3 - 1) + \mu_0(10x_1 + 8x_2 + 6x_3 - 7) \\ & + \mu_1x_1 + \mu_2x_2 + \mu_3x_3. \end{aligned}$$

Then, the KKT conditions are

$$
\begin{aligned}
L_{x_1} = -2x_1 - 0.5x_2 - 0.01x_3 + \lambda + 10\mu_0 + \mu_1 &= 0, & (1)\\
L_{x_2} = -0.5x_1 - x_2 + 0.01x_3 + \lambda + 8\mu_0 + \mu_2 &= 0, & (2)\\
L_{x_3} = -0.01x_1 + 0.01x_2 - 0.1x_3 + \lambda + 6\mu_0 + \mu_3 &= 0 & (3)
\end{aligned}
$$

and

$$L_\lambda = x_1 + x_2 + x_3 - 1 = 0 \tag{4}$$

and

$$L_{\mu_0} = 10x_1 + 8x_2 + 6x_3 - 7 \geq 0, \quad \mu_0 \geq 0, \quad \mu_0 L_{\mu_0} = 0$$

and

$$L_{\mu_i} = x_i \geq 0, \quad \mu_i \geq 0, \quad \mu_i x_i = 0, \quad i = 1, 2, 3.$$

Case 1: $10x_1 + 8x_2 + 6x_3 > 7$.

In this case, we always have $\mu_0 = 0$.

Case 1a: assume all x_i non-zero.

Then $\mu_1 = \mu_2 = \mu_3 = 0$.

Solving the equations (1)-(4),

$$x_1 = 0.0177, x_2 = 0.0887, x_3 = 0.8936, \lambda = 0.0886.$$

But we have

$$10x_1 + 8x_2 + 6x_3 = 6.2482 < 7.$$

This case will not happen.

Case 1b: assume $x_1 = 0$ and x_2 and x_3 non-zero.

Then we have $\mu_2 = \mu_3 = 0$.

Solving equations (2), (3) and (4),

$$x_2 = 0.0982, \quad x_3 = 0.9018, \quad \lambda = 0.0892.$$

Using equation (1),

$$\mu_1 = 2x_1 + 0.5x_2 + 0.03x_3 - \lambda = -0.0311$$

This case will not happen.

Case 1c: $x_2 = 0$, but $x_1, x_3 \neq 0$.
We have $\mu_1 = \mu_3 = 0$.
Solving eqs. (1), (3), (4)

$$x_1 = 0.0433, x_3 = 0.9567$$

and

$$\lambda = 0.0961.$$

Using equation (2),

$$\mu_2 = -0.084.$$

Case 1d: $x_3 = 0$, but $x_1, x_2 \neq 0$.
We have $\mu_1 = \mu_2 = 0$.
Solving eqs. (1), (2), (4)

$$x_1 = 0.25, x_3 = 0.75$$

and

$$\lambda = 0.875.$$

Using equation (3),

$$\mu_3 = -0.88.$$

Case 1e: $x_1 = x_2 = 0$. Then $x_3 = 1$ by (4), contradicts the first assumption.

Case 1f: $x_1 = x_3 = 0$. Then $x_2 = 1$ by (4). So, $\mu_2 = 0$.

Equation (2) implies $\lambda = 1$.

Equation (1) implies $\mu_1 = -0.5 < 0$.

Case 1g: $x_2 = x_3 = 0$. Then $x_1 = 1$ by (4). So, $\mu_1 = 0$.

Equation (1) implies $\lambda = 2$.

Equation (2) implies $\mu_2 = -\lambda + 0.5x_1 = -1.5$.

Finally, we conclude that Case 1 will not happen.

Case 2: $10x_1 + 8x_2 + 6x_3 = 7$.

Case 2a: assume all x_i non-zero.

Then $\mu_1 = \mu_2 = \mu_3 = 0$.

Solving this together with the first 4 equations,

$$x_1 = 0.1659, x_2 = 0.1683, x_3 = 0.6659, \lambda = -0.4674, \mu_0 = 0.0890.$$

Good!

Case 2b: assume $x_1 = 0$, but x_2, x_3 non-zero.

Solving equations (2), (3) and (4),

$$x_2 = 0.5, \quad x_3 = 0.5, \quad \lambda = -1.305, \quad \mu_0 = 0.225.$$

Equation (1) gives $\mu_1 = -0.69$. Wrong!

Case 2c: assume $x_2 = 0$, but x_1, x_3 non-zero.

Solving equations (1), (3) and (4),

$$x_1 = 0.25, \quad x_3 = 0.75, \quad \lambda = -0.5675, \quad \mu_0 = 0.1075.$$

From equation (2), we have $\mu_2 = -0.175$. Wrong!

Case 2d: assume $x_3 = 0$, but x_1, x_2 non-zero.

Solving equations (1), (2) and (4),

$$x_1 = -0.5 < 0.$$

Case 2e: $x_2 = x_3 = 0$, $x_1 \neq 0$, ...

Case 2f: $x_1 = x_3 = 0$, $x_2 \neq 0$, ...

Case 2g: $x_1 = x_2 = 0$, $x_3 \neq 0$, ...

We see that all KKT conditions are satisfied for Case 2a.

In conclusion, the solution is:

- invest 16.59% of your money in stocks;
- invest 16.83% of your money in bonds;
- invest 66.59% of your money in money markets;
- your return is 7%;
- the risk is 0.0778.

More words about optimization

- Optimization is not solely an applied math subject.
- Even with today's computing power, solving nonlinear optimization is still a challenge in many areas.
- General algorithms for nonlinear optimization may not be good, background knowledge is crucial.

Disclaimer

All figures, tables, and data appearing in the slides are only used for teaching under guidelines of Fair Use.