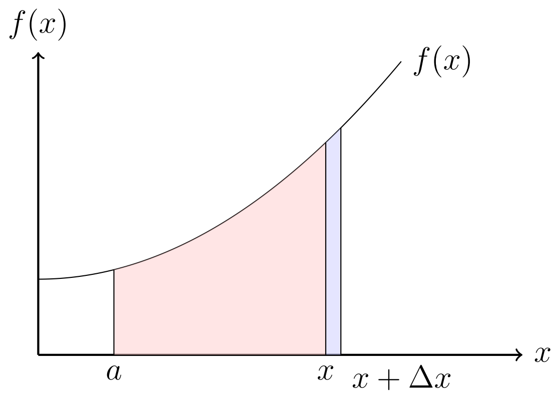
Topic 8: Integrals[1]

Fundamental Theorem of Calculus

Due to time constraints, we will skip the formal definition of integrability and focus instead on applications.

Denote the integral of a function f from a to x as $F(x)$: $F(x) = \int_a^x (t)dt$. From the diagram below, observe that when x increases by a small amount Δx, the change in the accumulated area, given by $F(x + \Delta x) - F(x)$, is approximately equal to the function value at x.



Theorem 1: First Fundamental Theorem of Calculus

Let f be a Riemann integrable function on $[a, b]$. For $a \leq x \leq b$, put

$$F(x) = \int_a^x f(t)\ dt.$$

Then, F is (uniformly) continuous on $[a, b]$.

Furthermore, if f is continuous at a point x_0 of $[a, b]$, then F is differentiable at x_0, and

$$F'(x_0) = f(x_0).$$

[1]Instructors: Camilo Abbate and Sofia Olguin. This note was prepared for the 2025 UCSB Math Camp for Ph.D. students in economics. It incorporates materials from previous instructors, including Shu-Chen Tsao, ChienHsun Lin, and Sarah Robinson.

The First Fundamental Theorem of Calculus suggests that integration and differentiation are inverse operations. For this reason, the integral of f is sometimes called the **antiderivative** of f.

Theorem 2: Second Fundamental Theorem of Calculus

Let f be a Riemann integrable function on $[a, b]$. Suppose there is a differentiable function F on $[a, b]$ such that $F' = f$, then

$$\int_a^b f(x)\ dx = F(b) - F(a).$$

Properties of Integrals

The following are some rules for integrals. Assume all functions are Riemann integrable.

Proposition 1

1. Linearity: $\int_a^b f(x) + g(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$.

2. Bounds: Let m be the lower bound of $f(x)$ for all $x \in [a, b]$, and let M be the upper bound of $f(x)$ for all $x \in [a, b]$. Then, $m(b - a) \leq \int_a^b f(x)dx \leq M(b - a)$.

3. Inequalities between functions: If $f(x) \leq g(x)$ for each $x \in [a, b]$, then $\int_a^b f(x)dx \leq \int_a^b g(x)dx$.

Example 1

Suppose a company expects to receive a continuous income stream at a rate of $R(t) = 2000e^{0.05t}$ dollars per year, where t is the time in years.

Calculate the present value of this income stream over the next 10 years if the discount rate is 2 percent.

One useful theorem that we learned in freshmen calculus is the integration by parts.

Theorem 3: Integration by parts

Suppose F and G are differentiable on $[a, b]$, and f and g are Riemann integrable. Then,

$$\int_a^b F(x)g(x)\ dx = (F(b)G(b) - F(a)G(a)) - \int_a^b f(x)G(x)\ dx$$

Example 2

Evaluate the following integral

$$\int_a^b xe^x\ dx.$$

The first step for integration by parts is to determine the parts.

Let $F(x) = x$, and $G(x) = e^x$. Then, $f(x) = F'(x) = 1$, and $g(x) = g'(x) = e^x$.

Hence, we can replace our integral by

$$\begin{aligned}\int_a^b xe^x dx &= be^b - ae^a - \int_a^b e^x dx \\ &= be^b - ae^a - \left[e^x\right|_a^b \\ &= be^b - ae^a - \left(e^b - e^a\right) = (b-1)e^b - (a-1)e^a.\end{aligned}$$

Leibniz rule

Theorem 4

Suppose $f(x,t)$ is continuously differentiable with respect to t, and $a(t), b(t)$ are differentiable.[a]

Then,

$$\frac{d}{dt}\left(\int_{a(t)}^{b(t)} f(x,t)\,dx\right) = \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(x,t)\,dx + f(b(t),t)\cdot b'(t) - f(a(t),t)\cdot a'(t)$$

[a] Continuously differentiable: the derivative exists and is continuous.

Example 3

Evaluate the following integral

$$\frac{d}{dt}\int_0^{t^2} e^{xt}\,dx.$$

Useful inequalities for integrals

This section revisits four useful inequalities: Jensen's, Cauchy-Schwarz, Hölder's, and Minkowski. Although we originally defined them in algebraic or metric space settings, each of these inequalities extends naturally to integrals.

These inequalities appear frequently in applications during the first year, particularly for econometrics courses.

Theorem 5: Jensen's Inequality in $\mathbb{R}^n$

Consider any $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^n$, and any $\lambda_1, ..., \lambda_n \in [0,1]$ such that $\sum_{i=1}^n \lambda_i = 1$. This forms a convex combination $\sum_{i=1}^n \lambda_i \mathbf{x}_i$. Suppose f is convex.

Jensen's inequality states:

$$f\left(\sum_{i=1}^n \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^n \lambda_i f(\mathbf{x}_i)$$

Proposition 2 (Jensen's Inequality in Integral Form)

Let g be a convex function and f a measurable function such that both $g(f(x))$ and $f(x)$ are integrable.

Jensen's inequality states:

$$g\left(\int_X f(x)\, d\mu(x)\right) \leq \int_X g(f(x))\, d\mu(x).$$

where $\mu(x)$ is a measure (e.g., probability measure) on the space X.

The statistical version of Jensen's inequality is

$$h(\mathbb{E}[X]) \leq \mathbb{E}[h(X)]$$

where h is a convex function, X is a random variable, and $\mathbb{E}[\cdot]$ is the expectation.

Theorem 6: Cauchy-Schwarz Inequality in $\mathbb{R}^n$

Consider any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$. The Cauchy-Schwarz inequality states:

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right)$$

Proposition 3 (Cauchy-Schwarz Inequality in Integral Form)

The Cauchy-Schwarz inequality states:

$$\left(\int_X f(x)g(x)\, d\mu(x)\right)^2 \leq \left(\int_X |f(x)|^2\, d\mu(x)\right)\left(\int_X |g(x)|^2\, d\mu(x)\right).$$

The statistical version of the Cauchy-Schwarz inequality is

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

Theorem 7: Hölder's Inequality in $\mathbb{R}^n$

Consider any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$. The Hölder's inequality states:

$$\sum_{i=1}^{n} |x_i y_i| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}}$$

for all $p, q \in [1, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Proposition 4 (Hölder's Inequality Inequality in Integral Form)

Hölder's inequality states:

$$\int_X |f(x)g(x)| \, d\mu(x) \leq \left(\int_X |f(x)|^p \, d\mu(x)\right)^{1/p} \left(\int_X |g(x)|^q \, d\mu(x)\right)^{1/q}.$$

for all $p, q \in [1, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$.

The statistical version of Hölder's inequality is

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^q]^{\frac{1}{q}}$$

Theorem 8: Minkowski inequality $\mathbb{R}^n$

Consider any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$. The Minkowski inequality states:

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}} \quad \text{for all } p \in [1, \infty).$$

Proposition 5 (Minkowski inequality in Integral Form)

The Minkowski inequality states:

$$\left(\int_X |f(x) + g(x)|^p \, d\mu(x)\right)^{1/p} \leq \left(\int_X |f(x)|^p \, d\mu(x)\right)^{1/p} + \left(\int_X |g(x)|^p \, d\mu(x)\right)^{1/p}.$$

The statistical version of the Minkowski inequality is

$$(\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^p]^{\frac{1}{p}}$$

Topic 9: Optimization[1]

Optimization Problems

Optimization problems are fundamental to economics because they provide a framework for modeling and analyzing how agents make decisions under constraints.

To further motivate this topic, consider a classic consumer choice problem. Suppose an individual consumes two goods x_1 and x_2, with the following utility function and budget constraint:

$$u(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha} \qquad p_1 x_1 + p_2 x_2 \leq m \qquad x_1, x_2 \in \mathbb{R}_+.$$

How can we formally express the problem of finding the optimal consumption (x_1, x_2) to maximize utility?

Definition 1

Let $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}$, and $h : X \to \mathbb{R}$. Also let $(\alpha, \beta, \gamma, \theta) \in \Theta$. Then the following expression

$$\max_{x \in D(\theta)} f(x; \alpha) \qquad \text{subject to } g(x; \beta) = 0,\ h(x; \gamma) \geq 0.$$

is a **maximization** problem of x.

- f is the **objective function**.
- x is the **choice variable(s)**. $D(\theta)$ is the **choice set**.
- g is the **equality constraint**. h is the **inequality constraint**.
- α, β, γ, θ are **parameters**. Θ is the **parameter space**.

Then we define the **solution** of a maximization problem.

[1]Instructors: Camilo Abbate and Sofia Olguin. This note was prepared for the 2025 UCSB Math Camp for Ph.D. students in economics. It incorporates materials from previous instructors, including Shu-Chen Tsao, ChienHsun Lin, and Sarah Robinson.

Definition 2

Let the optimization problem $\mathcal{D}$ be

$$\max_{x \in D(\theta)} f(x; \alpha) \qquad \text{subject to } g(x; \beta) = 0,\ h(x; \gamma) \geq 0.$$

Suppose there exists $x^* \in D(\theta)$, $g(x^*; \beta) = 0$, and $h(x^*; \gamma) \geq 0$ such that

$$f(x^*) \geq f(x) \text{ for every } x \text{ such that } g(x; \beta) = 0,\ h(x; \gamma) \geq 0.$$

Then x^* is called the a **solution** of $\mathcal{D}$.

We can write $\boldsymbol{x}^* : \Theta \rightrightarrows X$ as a correspondence such that that

$$\boldsymbol{x}^*(\alpha, \beta, \gamma, \theta) = \arg \max_{x \in D(\theta)} f(x; \alpha) \qquad \text{subject to } g(x; \beta) = 0,\ h(x; \gamma) \geq 0.$$

That is, $\boldsymbol{x}^* = \{x \in X | x^* \text{ is a solution of } \mathcal{D}\}$. Then $\boldsymbol{x}^*$ is called the **solution set** of $\mathcal{D}$.

The optimization problem at the very beginning can be expressed as

$$\max_{x_1, x_2 \in \mathbb{R}_+} x_1^{\alpha} x_2^{1-\alpha} \qquad \text{subject to } p_1 x_1 + p_2 x_2 \leq m.$$

Unconstrained Optimization Problems

Let us start by solving optimization problems without constraints.

Example 1

Let $f(x, y) = |x - y|$. $x \in \{0, 1, \ldots, n\}$, $y \in \{0, 1, \ldots, n\}$, where $n \in \mathbb{N}$. Find the solution(s) (x^*, y^*) that maximizes f.

Example 2

Let $f(x, y) = |x - y|$. The constraint is $x + y \leq 5$. $x, y \in \mathbb{R}_+$ (including 0). Find the solution(s) (x^*, y^*) that maximizes f.

Example 3

Let $f(x,y) = |x-y|$. The constraint is $x+y<5$. $x,y \in \mathbb{R}_+$ (including 0). Does the solution that maximizes f exist?

Given the solution and the parameters, the value of the function is then called the **value function**.

$$V(\alpha,\beta,\gamma,\theta) = f(x^*(\alpha,\beta,\gamma,\theta),\alpha).$$

In Example 1, the solution set is

$$(x^*(n), y^*(n)) = \{(0,n),(n,0)|n \in \mathbb{N}\}.$$

Plugging in the solution, we can find the value function

$$V(n) = f(x^*(n), y^*(n)) = |n-0| = n.$$

You can find that given different values of the parameter n, the optimal solutions and values may be different.

First Order and Second Order Conditions

We now consider the convenient case where the objective function is differentiable and there are no constraints. The first order condition serves as our starting point, allowing us to identify solution candidates.

Theorem 1

Let $f:\mathbb{R}\to\mathbb{R}$ be defined on $[a,b]$. If f has a local extreme at $x \in (a,b)$, and if $f'(x)$ exists, then $f'(x) = 0$.

Notice that the first order condition can only detect the **interior solutions** of the optimization function. If the solutions are on the boundaries, the solutions are called the **corner solutions**.

To identify the extreme value is a maximum or a minimum, we can further apply the **second order condition**.

Theorem 2

Let $f : \mathbb{R} \to \mathbb{R}$ be defined on $[a, b]$, f is twice differentiable, and $f'(x) = 0$. If $x \in (a, b)$ is a local maximum of f, then $f''(x) \leq 0$; if $x \in (a, b)$ is a local minimum of f, then $f''(x) \geq 0$.

Instead of a local extreme, we are looking for a **global extreme** such that $f(x^*) \geq f(x)$ for every $x \in X$.

Summary: To find the **global extreme** such that $f(x^*) \geq f(x)$ for every $x \in X$, we need to check the following (say, for maximization problem):

- Find all x such that $f'(x) = 0$ and $f''(x) \leq 0$. Then, evaluate $f(c)$.
- Evaluate $f(x)$ at the corner(s) of X.
- Compare all $f(x)$ such that x satisfy either of the two conditions.

Example 4

Find the solution to the following optimization problems.

1. $\max_{x \in [-1,2]} -x^4 + 2x^3 + 1$
2. $\min_{x \in [-1,2]} -x^4 + 2x^3 + 1$

We can also extend our scope to multivariate functions.

Theorem 3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined on $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$. If f has a local extreme at $\mathbf{x} \in E \subset I$ for an open set E, and if $\nabla f(\mathbf{x})$ exists, then $\nabla f(\mathbf{x}) = \mathbf{0}$.

We can extend the second order condition to multivariate functions. Recall that the Hessian matrix contains all second order partial derivatives.

Theorem 4

Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined on $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$. Let f be twice differentiable and $\nabla f(\mathbf{x}) = \mathbf{0}$. Let $E \in I$ be an open set, and $\mathbf{H} = D(\nabla f)$ be the Hessian matrix.

Then, if $\mathbf{x} \in E$ is a local maximum of f, then $\mathbf{H}$ is negative semidefinite; if $\mathbf{x} \in E$ is a local minimum of f, then $\mathbf{H}$ is positive semidefinite.

Summary: To find the **global extreme** such that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for every $\mathbf{x} \in E$, we need to check the following facts (say, for maximization problem).

- Find all $\mathbf{x}$ such that $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\mathbf{H}$ is negative semidefinite. Then, evaluate $f(\mathbf{x})$.
- Evaluate $f(\mathbf{x})$ at the corner(s) of E.
- Compare all $f(\mathbf{x})$ such that $\mathbf{x}$ satisfy either of the two conditions.

Example 5

Identify the extrema for the function

$$f(x, y) = 8x^3 + 2xy - 3x^2 + y^2 + 1.$$

Envelope Theorem

Consider the profit maximization problem:

$$\max_{K>0, L>0} \pi(K, L; p, w, r) = \max_{K>0, L>0} p\left[\log(K) + \log(L)\right] - rK - wL.$$

You can find the solution

$$K^*(p, w, r) = \frac{p}{r}, \quad L^*(p, w, r) = \frac{p}{w}$$

and the value function

$$V_\pi(p, w, r) = p\left[\log\left(\frac{p}{r}\right) + \log\left(\frac{p}{w}\right)\right] - 2p.$$

We are often interested in the extra value of a marginal increase in price p. In math, this term is:

$$\frac{\mathrm{d}V_\pi}{\mathrm{d}p} \equiv \frac{\mathrm{d}}{\mathrm{d}p}\pi(K^*, L^*, p; r, w)$$

You may notice that this total derivative can be difficult to calculate, since the optimal choice $K^*(p,w,r) = \frac{p}{r}$ and $L^*(p,w,r) = \frac{p}{w}$ change as well when p changes. Luckily, we have the envelope theorem.

In short, the envelope theorem states that the total derivative $\frac{\mathrm{d}}{\mathrm{d}p}\pi(K^*, L^*, p; r, w)$ is equivalent to the partial derivative $\frac{\partial \pi}{\partial p}(K^*, L^*, p; r, w)$, which is often easier to calculate. To verify this, we can evaluate the derivatives of K^*, L^*, and V_π with respect to p.

$$\frac{\mathrm{d}K^*}{\mathrm{d}p} = \frac{1}{r}, \quad \frac{\mathrm{d}L^*}{\mathrm{d}p} = \frac{1}{w}, \quad \frac{\mathrm{d}V_\pi}{\mathrm{d}p} = \log\left(\frac{p}{r}\right) + \log\left(\frac{p}{w}\right).$$

Let's see a formal proof of the envelope theorem for this case. Note that the value function is derived from the objective function using the optimized choices.

$$\begin{aligned}
\mathrm{d}V_\pi \equiv & \mathrm{d}\pi(K^*, L^*, p; r, w) \\
= & D_p\pi(K^*, L^*, p; r, w)\mathrm{d}p \\
& + D_K\pi(K^*, L^*; p, r, w)\mathrm{d}K + D_L\pi(K^*, L^*; p, r, w)\mathrm{d}L.
\end{aligned}$$

Since now we are interested in how p will change the whole value, we also include p as a variable. Hence

$$\begin{aligned}
\frac{\mathrm{d}V_\pi}{\mathrm{d}p} = & D_p\pi(K^*, L^*, p; r, w) && \text{(direct effect)} \\
& + D_K\pi(K^*, L^*; p, r, w)\frac{\mathrm{d}K}{\mathrm{d}p} + D_L\pi(K^*, L^*; p, r, w)\frac{\mathrm{d}L}{\mathrm{d}p}. && \text{(indirect effect)}
\end{aligned}$$

However, we know that K^* and L^* are (local) maxima. So, according to the first order condition:

$$D_K\pi(K^*, L^*; p, r, w) = D_L\pi(K^*, L^*; p, r, w) = 0.$$

Therefore,

$$\frac{\mathrm{d}V_\pi}{\mathrm{d}p} \equiv \frac{\mathrm{d}}{\mathrm{d}p}\pi(K^*, L^*, p; r, w) = \frac{\partial \pi}{\partial p}(K^*, L^*, p; r, w).$$

In other words, *under optimality*, since the partial derivative with respect to the choice variables are all equal to zero, we don't need to worry about the indirect effects of p on optimal consumption of K and L. Hence, we can simply evaluate the comparative statics with the direct effect at the optimal choices. This result is called the **envelope theorem**.

Theorem 5: Envelope Theorem: unconstrained version

Consider $f : \mathbb{R}^n \to \mathbb{R}$, $\theta \in \Theta$ and a maximization problem $\max_x f(x;\theta)$, and let x^* be the solution function. If f is differentiable and if x^* is continuously differentiable with respect to θ at $\overline{\theta}$, then for the value function $v : \Theta \to \mathbb{R}$,

$$\frac{\mathrm{d}v}{\mathrm{d}\theta} \equiv \frac{\mathrm{d}f}{\mathrm{d}\theta}(x^*;\theta) = \frac{\partial f}{\partial \theta}(x^*;\theta)$$

at the specific $\overline{\theta}$.

Optimization Problems with Equality Constraints

Let us now turn to the optimization with constraints. Consider, again, the following utility function and the budget constraint:

$$u(x_1, x_2) = x_1^\alpha x_2^{1-\alpha} \qquad p_1x_1 + p_2x_2 = m.$$

We have an equality constraint to this question. One way to solve this is to replace x_2 with

$$x_2 = \frac{m - p_1x_1}{p_2}$$

and then plug it back to the objective function, so that the optimization problem becomes unconstrained again.

However, if the constraint is like

$$\exp\left(\frac{\sqrt{x_1^3 + x_2^{1/\alpha}}}{2\pi(x_2 - \mu)^2}\right) = x_1\sigma^2$$

then the replacement method does not seem to be very efficient. Hence, we introduce an extremely powerful tool that can be applied to almost *all* situations: Lagrangian function.

Definition 3

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Also let $(\alpha, \beta) \in \Theta$. Then for the following optimization problem

$$\max_{\mathbf{x}} f(\mathbf{x}; \alpha) \qquad \text{subject to } g(\mathbf{x}; \beta) = 0.$$

We can write a **Lagrangian function**

$$\mathcal{L}(\mathbf{x}, \lambda; \alpha, \beta, \theta) = f(\mathbf{x}; \alpha) + \lambda g(\mathbf{x}; \beta),$$

where λ is called the **Lagrangian multiplier**.

The following theorem identifies the extreme value under the constrained optimization problem.

Theorem 6: Lagrange's theorem

Let f and g are continuously differentiable over some $D \subset \mathbb{R}^n$. Consider an optimization problem

$$\max_{\mathbf{x}} \ f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) = c.$$

Let the **Lagrangian function** be

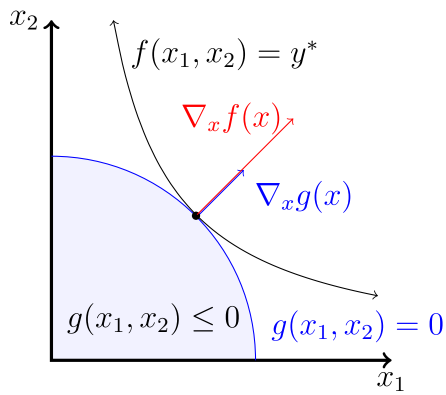
$$\mathcal{L}(\mathbf{x}, \lambda; \alpha, \beta, c) = f(\mathbf{x}; \alpha) + \lambda(c - g(\mathbf{x}; \beta)).$$

Suppose $\mathbf{x}^*$ is an interior point of D, and $\mathbf{x}^*$ solves the optimization problem. Then, there exists $\lambda^* \in R$ such that for every $i = 1, \ldots, n$,

$$\frac{\partial \mathcal{L}}{\partial x_i}(\mathbf{x}^*, \lambda^*) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = 0.$$

To some extent, Lagrangian function transforms the constrained optimization into an unconstrained problem, so we can simply evaluate the optimization problem of the Lagrangian.

To see the reasoning behind the theorem, first notice that when the constraint is satisfied, the solution that optimizes the Lagrangian also optimizes the original function. Also, if we look at the level curves on the choice space X, the gradient vector represents the *normal vector* at some specific point.



When the objective function is optimized given some constraint, the function and the constraint must be tangent with each other on the choice space. That is, the normal vectors of the f and g must be linear dependent at that specific point. Therefore, there exists a scalar λ such that $\nabla_x f(x^*) = \lambda \nabla_x g(x^*)$.

The Lagrange multiplier λ is sometimes interpreted as the **shadow price** of the constraint. Since

$$\mathrm{d}f(x) = \nabla_x f(x^*) \cdot \mathrm{d}x = \lambda \nabla_x g(x) \cdot \mathrm{d}x = \lambda \mathrm{d}g(x),$$

one may interpret λ as the marginal value of the constraint. For example, suppose that $g(x) = m - p \cdot x$ is the budget constraint in the consumption maximization problem, and $f(x)$ is the utility function. Then $\mathrm{d}g$ is the marginal change in the budget constraint, namely the income, $\mathrm{d}m$. Then when the income m increases by $\mathrm{d}m$, the total utility increases by $\lambda \mathrm{d}g = \lambda \mathrm{d}m$ units.

Another implication is to view Lagrangian multiplier as a dual variable. Suppose, instead of solving the optimization of f, we are now interested in the optimization of g. (For example, in the scenario of utility maximization problem, now we are fixing the utility at a certain level and minimizing the expenditure). Then we can use the similar construction and find

$$\gamma \nabla f(x^*) = \nabla g(x)$$

and the Lagrange multiplier $\gamma = 1/\lambda$. You may (or may not) learn more about the duality problem during the first quarter microeconomic theory.

We provide the sufficient condition for the solutions to constrained optimization problem for the special case that $f : \mathbb{R}^2 \to \mathbb{R}$ with a single constraint.

Theorem 7: Sufficient condition of Constrained Optimization

Suppose the optimization problem is:

$$\max_{x_1,x_2} f(x_1,x_2) \text{ subject to } c = g(x_1,x_2)$$

or

$$\min_{x_1,x_2} f(x_1,x_2) \text{ subject to } c = g(x_1,x_2)$$

Denote the Lagrangian as $\mathcal{L}(x_1, x_2, \lambda)$.

Define the **bordered Hessian matrix**

$$\overline{\mathbf{H}} = \begin{bmatrix} \mathcal{L}_{\lambda\lambda} & \mathcal{L}_{\lambda 1} & \mathcal{L}_{\lambda 2} \\ \mathcal{L}_{1\lambda} & \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{2\lambda} & \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{L}_{\lambda K} & \mathcal{L}_{\lambda L} \\ \mathcal{L}_{K\lambda} & \mathcal{L}_{KK} & \mathcal{L}_{KL} \\ \mathcal{L}_{L\lambda} & \mathcal{L}_{LK} & \mathcal{L}_{LL} \end{bmatrix}.$$

Suppose $\nabla\mathcal{L}(x_1^*, x_2^*, \lambda^*) = 0$.

Then, if $\det(\overline{\mathbf{H}}) > 0$, then (x_1^*, x_2^*) solves the maximization problem; if $\det(\overline{\mathbf{H}}) < 0$, then (x_1^*, x_2^*) solves the minimization problem.

Example 6

Suppose the cost function is $C(K, L) = (rK + wL)^2$ $(r, w > 0)$, and the production function is $F(K, L) = KL$, where $\alpha \in (0, 1)$. Solve the cost minimization problem

$$\min_{K,L} C(K, L) \text{ subject to } F(K, L) = y$$

for some given level of y, and then find the total cost function $C(r, w, y)$.

Solution. First of all, we write the Lagrangian and solve the gradient.

$$\mathcal{L}(K, L, \lambda) = (rK + wL)^2 + \lambda(y - KL).$$

$$\frac{\partial \mathcal{L}}{\partial K} = 2r(rK + wL) - \lambda L = 0 \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial L} = 2w(rK + wL) - \lambda K = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = y - KL = 0 \tag{3}$$

From (1) and (2), we can solve

$$\lambda = \frac{2r(rK + wL)}{L} = \frac{2w(rK + wL)}{K} \;\Rightarrow\; rK^* = wL^*.$$

Then we can find

$$K^*L^* = K^* \cdot \left(\frac{r}{w}\right) K^* = y.$$

Hence

$$K^* = \sqrt{\frac{wy}{r}}, \quad L^* = \sqrt{\frac{ry}{w}}, \quad (\lambda^* = 4rw)$$

and

$$C(r, w, y) = C(K^*(r, w, y), L^*(r, w, y)) = \left(r\sqrt{\frac{wy}{r}} + w\sqrt{\frac{ry}{w}}\right)^2 = 4rwy.$$

One can also check the bordered Hessian for second order conditions

$$\begin{vmatrix} \mathcal{L}_{\lambda\lambda} & \mathcal{L}_{\lambda K} & \mathcal{L}_{\lambda L} \\ \mathcal{L}_{K\lambda} & \mathcal{L}_{KK} & \mathcal{L}_{KL} \\ \mathcal{L}_{L\lambda} & \mathcal{L}_{LK} & \mathcal{L}_{LL} \end{vmatrix} = \begin{vmatrix} 0 & -L & -K \\ -L & 2r^2 & 2wr - \lambda \\ -K & 2wr - \lambda & 2w^2 \end{vmatrix}$$

$$= L\left[-2w^2L + K(2wr - \lambda)\right] - K\left[-L(2wr - \lambda) + 2r^2K\right]$$

$$= -2w^2L^2 - 2r^2K^2 + 4wrKL - 2LK\lambda = -2(rK - wL)^2 - 2LK\lambda < 0.$$

Hence K^* and L^* solve the minimization problem. ■

For completeness, we provide the generalized version of Lagrange's theorem, where we have more than one constraint.

Definition 4

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g^k : \mathbb{R}^n \to \mathbb{R}$ for each $k = 1, \ldots, K$. Also let $(\alpha, \beta) \in \Theta$. Then for the following optimization problem

$$\max_{\mathbf{x}} f(\mathbf{x}; \alpha) \text{ subject to } c_1 = g^1(\mathbf{x}; \beta), \ldots, c_K = g^k(\mathbf{x}; \beta),$$

We can write a Lagrangian function

$$\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_K; \alpha, \beta, \mathbf{c}) = f(\mathbf{x}; \alpha) + \lambda_1[c_1 - g^1(\mathbf{x}; \beta)] + \cdots + \lambda_K[c_K - g^k(\mathbf{x}; \beta)],$$

where $\lambda_1, \ldots, \lambda_K \in \mathbb{R}$ are Lagrange multipliers.

Theorem 8: Lagrange's theorem: multiple constraints

Let f and g are continuously differentiable over some $D \subset \mathbb{R}^n$. Consider an optimization problem and the Lagrangian function above. Suppose $\mathbf{x}^*$ is an interior point of D, $\mathbf{x}^*$ solves the optimization problem, and $\nabla g^k(\mathbf{x}^*)$ is linear independent for each k. Then there exist unique $\lambda^* \in R^K$ such that for every $i = 1, \ldots, n$ and $k = 1, \ldots, K$,

$$\frac{\partial \mathcal{L}}{\partial x_i}(\mathbf{x}^*, \lambda^*) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda_k}(\mathbf{x}^*, \lambda^*) = 0.$$

Envelope Theorem—With Constraints

After solving the constrained maximization problem, we can also find the comparative statics for the value function. Specifically, consider the following maximization problem:

$$\max_{\mathbf{x}} f(\mathbf{x}; \theta) \text{ subject to } c = g(\mathbf{x}; \theta).$$

We can write the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda; \theta, c) = f(\mathbf{x}; \theta) + \lambda[c - g(\mathbf{x}; \theta)].$$

Then the first order conditions yields

$$\frac{\partial \mathcal{L}}{\partial x_i} = f_i(\mathbf{x}^*; \theta) - \lambda g_i(\mathbf{x}^*; \theta) = 0 \;\Rightarrow\; f_i(\mathbf{x}^*; \theta) = \lambda g_i(\mathbf{x}^*; \theta) \quad \forall i.$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = c - g(\mathbf{x}^*; \theta) = 0 \;\Rightarrow\; g(\mathbf{x}^*; \theta) = c.$$

Note that, similarly, the total differential

$$\begin{aligned}
\mathrm{d}V_f \equiv \mathrm{d}f(\mathbf{x}^*,\theta) =& f_\theta(\mathbf{x}^*,\theta)\mathrm{d}\theta + f_1(\mathbf{x}^*,\theta)\mathrm{d}x_1 + \cdots + f_n(\mathbf{x}^*,\theta)\mathrm{d}x_n \\
=& f_\theta[\mathbf{x}^*,\theta)\mathrm{d}\theta + \lambda[g_1(\mathbf{x}^*,\theta)\mathrm{d}x_1 + \cdots + g_n(\mathbf{x}^*,\theta)\mathrm{d}x_n] \\
=& f_\theta(\mathbf{x}^*,\theta)\mathrm{d}\theta + \lambda\left[\mathrm{d}g(\mathbf{x}^*,\theta) - g_\theta(\mathbf{x}^*,\theta)\mathrm{d}\theta\right].
\end{aligned}$$

<u>Question</u>: How can we derive the second line? How can we derive the third line?

Therefore,

$$\begin{aligned}
\frac{\mathrm{d}V_f}{\mathrm{d}\theta} &= f_\theta(\mathbf{x}^*,\theta) + \lambda\left[\frac{\mathrm{d}g}{\mathrm{d}\theta}(\mathbf{x}^*,\theta) - g_\theta(\mathbf{x}^*,\theta)\right] \\
&= f_\theta(\mathbf{x}^*,\theta) - \lambda g_\theta(\mathbf{x}^*,\theta) && \left(g(\mathbf{x}^*,\theta) = c \;\Rightarrow\; \frac{\mathrm{d}g}{\mathrm{d}\theta} = 0\right) \\
&= \frac{\partial\mathcal{L}}{\partial\theta}(\mathbf{x}^*,\lambda^*,\theta).
\end{aligned}$$

<u>Question</u>: How do we derive the last line?

In short, instead of taking the total derivatives, you can just take the partial derivative of the Lagrangian *and* evaluate at the optimal.

Theorem 9: Envelope Theorem: constrained version

Consider $f : \mathbb{R}^n \to \mathbb{R}$, $\theta \in \Theta$ and a maximization problem

$$\max_x f(x;\theta) \text{ subject to } \boldsymbol{g}(x;\theta) \geq 0.$$

Let $\mathcal{L}(x,\lambda;\theta)$ be the corresponding Lagrangian. If f is differentiable and if the solution x^* is differentiable at some parameter $\overline{\theta}$, then for the value function $v : \Theta \to \mathbb{R}$,

$$\frac{\mathrm{d}v}{\mathrm{d}\theta} = \frac{\partial\mathcal{L}}{\partial\theta}(\mathbf{x}^*,\lambda^*,\theta)$$

at the specific $\overline{\theta}$.

Following from the cost minimization example we solved,

$$\frac{\partial C}{\partial r}(r, w, y) = \frac{\partial \mathcal{L}}{\partial r}(\mathbf{x}^*, \lambda^*, r, w, y) = 2K^*(rK^* + wL^*) = 4wy$$

$$\frac{\partial C}{\partial w}(r, w, y) = \frac{\partial \mathcal{L}}{\partial w}(\mathbf{x}^*, \lambda^*, r, w, y) = 2L^*(rK^* + wL^*) = 4ry$$

$$\frac{\partial C}{\partial y}(r, w, y) = \frac{\partial \mathcal{L}}{\partial y}(\mathbf{x}^*, \lambda^*, r, w, y) = \lambda = 4rw$$

The envelope theorem for constrained optimization problems is very common in economics. We use this to prove several important theorems in consumer theory.

Optimization Problems with Inequality Constraints

Karush-Kuhn-Tucker Condition

Now we start looking at the optimization problem with inequality constraints. For example,

$$\max_{\mathbf{x} \geq \mathbf{0}} u(\mathbf{x}) \text{ subject to } \mathbf{p} \cdot \mathbf{x} \leq m.$$

In this case, the constraint may not be **binding**, *i.e.* the "equal" part of the constraint may not be applied to the problem. For example,

$$\max_{x \geq 0, y \geq 0} -\sqrt{(x-1)^2 + (y-1)^2} \text{ subject to } x + y \leq 5.$$

In this utility maximization problem, there is a global bliss point $(x^*, y^*) = (1, 1)$ that maximizes the utility, and $x^* + y^* = 2 < 5$ where the budget constraint is **slack**. To cope with this type of the issue, we introduce the following **Karush-Kuhn-Tucker conditions**.

Theorem 10: Constrained Optimization—inequality constraints

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g^k : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable over some $D \subset \mathbb{R}^n$ for each $k = 1, \ldots, K$. Then consider the following optimization problem

$$\max_{\mathbf{x}} f(\mathbf{x}; \alpha) \text{ subject to } c_1 \geq g^1(\mathbf{x}; \beta), \ldots, c_K \geq g^K(\mathbf{x}; \beta),$$

We can write the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_K; \alpha, \beta, \mathbf{c}) = f(\mathbf{x}; \alpha) + \lambda_1[c_1 - g^1(\mathbf{x}; \beta)] + \cdots + \lambda_K[c_K - g^K(\mathbf{x}; \beta)].$$

If $\mathbf{x}^*$ is an interior point of D which solves the optimization problem, and $\nabla_{\mathbf{x}} g^k(\mathbf{x}^*)$ is linear independent for each k. Then there exists unique $\lambda \in \mathbb{R}^K$ satisfies the following conditions:

1. $\frac{\partial \mathcal{L}}{\partial x_i} = 0$ for each $i = 1, \ldots, n$
2. $\frac{\partial \mathcal{L}}{\partial \lambda_k} \geq 0$ for each $k = 1, \ldots, K$
3. $\lambda_k \geq 0$ for each $k = 1, \ldots, K$
4. $\lambda_k[c_k - g^k(\mathbf{x})] = 0$ for each $k = 1, \ldots, K$.

Be **very careful** about how the Lagrangian is settled up, especially for the signs and the inequalities in the condition. Also note that this set of the conditions are solving for the maximizing solutions. For the minimization problem, you will need to switch the problem into a maximization problem by adding a negative sign to **objective function**.

The last condition is also called the **complementary slackness condition**. It is the product of the Lagrange multiplier and the constraint. If the constraint k is not binding, that is, is *slack*,

$$c_k - g^k(\mathbf{x}) > 0$$

then according to the complementary-slackness condition, the Lagrange multiplier λ_k must be equal to 0. In such a situation, it is equivalent to have a Lagrangian where we remove constraint k.

Example 7

Solve the following maximization problem.

$$\max_{x,y} xy \text{ subject to } m \geq p_x x + p_y y, \; x \geq 0, \; y \geq 0.$$

Solution. We first write down the Lagrangian, carefully,

$$\mathcal{L}(x, y, \lambda, \mu_x, \mu_y) = xy + \lambda(m - p_x x - p_y y) + \mu_x x + \mu_y y.$$

Note that we have three Lagrange multipliers, λ, μ_x, and μ_y. Then according to KKT conditions, there exists non-negative λ, μ_x, and μ_y such that

$$\frac{\partial \mathcal{L}}{\partial x} = y^* - \lambda p_x + \mu_x = 0 \qquad \frac{\partial \mathcal{L}}{\partial y} = x^* - \lambda p_y + \mu_y = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = m - p_x x^* - p_y y^* \geq 0 \qquad \frac{\partial \mathcal{L}}{\partial \mu_x} = x^* \geq 0 \qquad \frac{\partial \mathcal{L}}{\partial \mu_y} = y^* \geq 0$$

and the complementary slackness condition:

$$\lambda[m - p_x x^* - p_y y^*] = \mu_x x^* = \mu_y y^* = 0.$$

Now, we have $\frac{\partial \mathcal{L}}{\partial x} = 0$, $\frac{\partial \mathcal{L}}{\partial y} = 0$, $\frac{\partial \mathcal{L}}{\partial \lambda} \geq 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} \geq 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} \geq 0$. How many cases should we check? There will be $2^3 = 8$ cases. Let's check all the cases.

1. <u>Case 1</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $m - p_x x^* - p_y y^* = 0, x^* = 0, y^* = 0$, which leads to contradiction when $m, p_x, p_y > 0$.
2. <u>Case 2</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $\lambda = 0$, $x^* = 0$, $y^* = 0$. So, $u(x^*, y^*) = 0$.
3. <u>Case 3</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $m - p_x x^* - p_y y^* = 0$, $\mu_x = 0$, $y^* = 0$. So, $u(x^*, y^*) = 0$.
4. <u>Case 4</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   Then, $m - p_x x^* - p_y y^* = 0$, $x^* = 0$, $\mu_y = 0$. So, $u(x^*, y^*) = 0$.
5. <u>Case 5</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $y^* = 0$. So, $u(x^*, y^*) = 0$.
6. <u>Case 6</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   Then, $x^* = 0$. So, $u(x^*, y^*) = 0$.
7. <u>Case 7</u>: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   Then, $m - p_x x^* + p_y y^* = 0$. From the first order conditions we can derive

   $$\lambda = -\frac{y^*}{p_x} = -\frac{x^*}{p_y} \;\Rightarrow\; p_x x = p_y y.$$

   Hence

   $$m = p_x x^* + p_y y^* = 2p_x x^* = 2p_y y^* \;\Rightarrow\; x^* = \frac{m}{2p_x},\; y^* = \frac{m}{2p_y}.$$

Then $u(x^*, y^*) = \frac{m^2}{4p_x p_y} > 0$.

8. Case 8: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_x} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   Then, $\lambda = \mu_x = \mu_y = 0$. From the first order conditions, we can derive $x^* = y^* = 0$. So, $u(x^*, y^*) = 0$.

Summarizing the cases, we find that the maximum is reached when only the constraint $m - p_x x^* - p_y y^* = 0$ binds, and $x^* = \frac{m}{2p_x}$, $y^* = \frac{m}{2p_y}$. ■

Notice that when we include the non-negative constraints, we add the (non-negative) Lagrange multipliers. Thus, using the same example,

$$\frac{\partial \mathcal{L}}{\partial x} = y^* - \lambda p_x + \mu_x = 0 \;\Rightarrow\; y^* - \lambda p_x = -\mu_x \leq 0,$$

where the equality binds when $\mu_x = 0$, or when the condition $x > 0$ holds.

Therefore, sometimes people do not explicitly add the non-negative constraints, but instead express the first two FOCs as:

$$\frac{\partial \mathcal{L}}{\partial x} \leq 0, \qquad \text{equality holds when } x > 0,$$

$$\frac{\partial \mathcal{L}}{\partial y} \leq 0, \qquad \text{equality holds when } y > 0.$$

Let's look at a trickier example with quasilinear objective functions.

Example 8

Suppose $p_x, p_y, m > 0$. Solve

$$\max_{x \geq 0, y \geq 0} \log(x) + y \text{ subject to } p_x x + p_y y \leq m.$$

Solution. We analyze the problem first: note that $\log(\cdot)$ does not have non-positive elements in the domain, so the condition $x = 0$ never binds. In other words, $\mu_x = 0$.

Then, we can write the Lagrangian and the KKT conditions.

$$\mathcal{L}(x, y, \lambda, \mu_y) = \log(x) + y + \lambda(m - p_x x - p_y y) + \mu_y y.$$

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{1}{x^*} - \lambda p_x = 0, \qquad \frac{\partial \mathcal{L}}{\partial y} = 1 - \lambda p_y + \mu_y = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = m - p_x x^* - p_y y^* \geq 0, \qquad \frac{\partial \mathcal{L}}{\partial \mu_y} = y^* \geq 0, \quad \lambda \geq 0, \ \mu_y \geq 0$$

$$\lambda(m - p_x x^* - p_y y^*) = 0, \qquad \mu_y y = 0.$$

Now consider the following cases.

1. Case 1: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $y^* = 0$. From the budget constraint, $x^* = \frac{m}{p_x}$, and $\lambda = \frac{1}{m}$, $\mu_y = \frac{p_y}{m} - 1$. Since $\mu_y \geq 0$, this case holds if $m \leq p_y$.
2. Case 2: $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   In this case, $\mu_y = 0$, so $\lambda = \frac{1}{p_y}$ and $x = \frac{p_y}{p_x}$. By the budget constraint, $y = \frac{m - p_y}{p_y} > 0$, and this case holds if $m > p_y$.
3. Case 3: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} = 0$.
   Then, $\lambda = 0$ and $y^* = 0$. This implies $\frac{\partial \mathcal{L}}{\partial x} = \frac{1}{x^*} = 0$. This is not possible for all $x \geq 0$.
4. Case 4: $\frac{\partial \mathcal{L}}{\partial \lambda} > 0$, $\frac{\partial \mathcal{L}}{\partial \mu_y} > 0$.
   Then, $\lambda = 0$. This implies $\frac{\partial \mathcal{L}}{\partial x} = \frac{1}{x^*} = 0$. This is not possible for all $x \geq 0$.

Therefore, we can summarize the solution as follows:

$$(x^*, y^*) = \begin{cases} \left(\frac{m}{p_x}, 0\right) & \text{if } m \leq p_y \\ \left(\frac{p_y}{p_x}, \frac{m - p_y}{p_y}\right) & \text{if } m > p_y \end{cases}.$$

■

Example 9

Suppose $p_x, p_y, u > 0$. Solve

$$\min_{x \geq 0, y \geq 0} p_x x + p_y y \text{ subject to } x + y \geq u.$$

For the minimization problem, we firstly reform the question into a maximization problem by switching the objective function:

$$\max_{x\geq 0, y\geq 0} -(p_x x + p_y y) \text{ subject to } x + y \geq u.$$

Then for $\lambda, \mu_x, \mu_y \geq 0$, the Lagrangian yields

$$\mathcal{L}(x, y, \lambda, \mu_x, \mu_y) = -(p_x x + p_y y) + \lambda(x + y - u) + \mu_x x + \mu_y y,$$

The rest is for your homework practice.

Sometimes, economic intuition is helpful for maximization problem. In this question, we are minimizing the expenditure spent on buying x and y, with the total units of x and y need to be no less than u.

Therefore, the solution to this question should looks like

$$(x^*, y^*) = \begin{cases} (0, u) & \text{if } p_x > p_y \\ (u, 0) & \text{if } p_x < p_y \end{cases}.$$

and when $p_x = p_y$, $(x^*, y^*) = (u - k, k)$, where $0 \leq k \leq u$. You can verify that this is indeed the solution.

The Sufficiency of Karush-Kuhn-Tucker Conditions*

As frequently emphasized, the first-order conditions derived from KKT framework are necessary conditions for optimal solutions , but not sufficient. That is, satisfying the KKT conditions does not guarantee that a candidate point is an optimal solution.

To illustrate, recall Example 7. In that case, the candidate solution $x^* = y^* = 0$ satisfies all KKT conditions, yet fails to achieve the maximum. The reason lies in the fact that the objective function, $f(x, y) = xy$, is not a concave function.

Below, we present two theorems that provide sufficient conditions under which KKT solutions are guaranteed to be optimal.

Theorem 11: Kuhn-Tucker Sufficiency Theorem

Consider the constrained maximization problem

$$\max_{\mathbf{x}\geq 0} f(x) \text{ subject to } g^1(\mathbf{x}), \ldots, g^K(\mathbf{x}).$$

Suppose the following conditions hold:

1. $f(\cdot)$ is differentiable and concave for $\mathbf{x} \geq 0$
2. $g^k(\cdot)$ is differentiable and convex for $\mathbf{x} \geq 0$ for each k
3. $\mathbf{x}^*$ satisfies the KKT maximum conditions.

Then $\mathbf{x}^*$ solves the maximization problem.

However, in practice, the concavity of f or the convexity of the constraints g^k may not hold. Fortunately, the following result provides a more flexible set of sufficient conditions that relax the concavity/convexity requirements.

Theorem 12: Arrow-Enthoven Sufficiency Theorem

Consider the maximization problem

$$\max_{\mathbf{x}\geq 0} f(x) \text{ subject to } g^1(\mathbf{x}), \ldots, g^K(\mathbf{x}).$$

Suppose all of the following conditions hold:

1. $f(\cdot)$ is differentiable and quasiconcave for $\mathbf{x} \geq 0$
2. $g^k(\cdot)$ is differentiable and quasiconvex for $\mathbf{x} \geq 0$ for each k
3. $\mathbf{x}^*$ satisfies the KKT maximum conditions.
4. any one of the following conditions is satisfied:
   (a) $f_j(\mathbf{x}^*) < 0$ for at least one x_j.
   (b) $f_j(\mathbf{x}^*) > 0$ for some x_j where $x_j^* > 0$ without violating any constraints
   (c) $\nabla f(\mathbf{x}^*) \neq 0$, and the second derivative exists at $(\mathbf{x}^*)$
   (d) $f(\cdot)$ is concave.

Then $\mathbf{x}^*$ solves the maximization problem.

Fortunately, most optimization problems encountered in practice in economics meet at least one of these sufficient conditions.

Hamiltonian and Optimal Control*

Note: This optional section covers differential equations.

Consider the following optimization problem.

$$\max_{c(t), t \geq 0} \int_0^T F(k) e^{-\rho t} \mathrm{d}t \quad \text{subject to } \dot{k}(t) = -c(t) - nk(t), \ k(0) = k_0$$

The optimization incorporates the integration over an interval of continuous time. The goal is to choose a series of $c(t)$ to maximize the time-discounted objective function, given the motion of capital $\dot{k}$ as the constraint. The "$\beta(t) \equiv e^{-\rho t}$" is called the **discounting factor**, where you can easily see that $\frac{\dot{\beta}}{\beta} = -\rho$. Hence this model captures the maximization problem at the end period ($t = T$) with a constant discount rate over time.

This type of maximization problem is called the **optimal control** problem. We formalize the optimal control problem as follows.

Definition 5 (Optimal control problem)

Let the optimal control problem be

$$\max_{x(t)} \int_0^T f(x(t), y(t), t) \mathrm{d}t \quad \text{subject to } \dot{y}(t) = g(x(t), y(t), t),$$

where f is integrable and satisfies necessary regularity conditions. $x(t)$ is called the **control variable**, and $y(t)$ is called the **state variable**. The solution of the optimal control problem is a function $x^* : [0, T] \to \mathbb{R}^n$.

In other words, we switch the control variable to adjust the state variable so that we can reach the optimality, and the "adjustment" is given by the motion function of the states. In the beginning example, $c(t)$ is the choice variable. and $k(t)$ is the state variable.

It seems intuitive to use the Lagrangian to solve the problem.

$$\mathcal{L} = \int_0^T f(x(t), y(t), t) - \lambda(t) \left[\dot{y}(t) - g(x(t), y(t), t)\right] \mathrm{d}t.$$

Note that the Lagrange multiplier, $\lambda(t)$, is also a function, as the constraint is a function over time.

As of Lagrangian, we will take the partial derivatives with respect to the variables x, y and λ. However, it is not so intuitive to "take derivatives" with respect to functions, $x(t)$, $y(t)$ and $\lambda(t)$. Instead, we make some adjustment to it. Note that by integration by part,

$$\int_0^T \lambda(t)\dot{y}(t)\mathrm{d}t = \lambda(t)y(t)\Big|_0^T - \int_0^T \dot{\lambda}(t)y(t)\mathrm{d}t.$$

Hence we can rewrite the Lagrangian as follows

$$\mathcal{L} = \int_0^T \left[f(x(t), y(t), t) + \lambda(t)g(x(t), y(t), t) + \dot{\lambda}(t)y(t)\right] \mathrm{d}t - \lambda(t)y(t)\Big|_0^T.$$

We focus on the integrand of the Lagrangian.

$$\mathcal{I} = f(x, y, t) + \lambda g(x, y, t) + \dot{\lambda}(t)y.$$

If this integrand is optimized for every x and y given each t, then the integral is also optimized. Therefore we can have the following FOCs.

$$\begin{aligned}\frac{\partial \mathcal{I}}{\partial x} =& f_x(x, y, t) + \lambda g_x(x, y, t) = 0 \\ \frac{\partial \mathcal{I}}{\partial y} =& f_y(x, y, t) + \lambda g_y(x, y, t) + \dot{\lambda} = 0.\end{aligned}$$

It seems that the solution depends only on the marginal behavior of $f + \lambda g$. Thus we define $f + \lambda g$ as the **Hamiltonian function**.

Proposition 1 (Hamiltonian)

Let the optimal control problem be

$$\max_{x(t)} \int_0^T f(x(t), y(t), t)\mathrm{d}t \quad \text{subject to } \dot{y}(t) = g(x(t), y(t), t),$$

where f is integrable and satisfies necessary regularity conditions. Define the **Hamiltonian function** as

$$\mathcal{H}(x, y, \lambda) = f(x, y, t) + \lambda g(x, y, t).$$

Then the solution of the optimal control problem satisfies the following first order conditions:

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial x} &= 0 \\ \frac{\partial \mathcal{H}}{\partial y} &= -\dot{\lambda} \\ \frac{\partial \mathcal{H}}{\partial \lambda} &= g(x, y)(= \dot{y}). \end{aligned}$$

λ is sometimes called the **costate**.

Example 10

Consider the *golden rule* maximization problem

$$\max_s \int_0^T (1-s)f(k(t))e^{-\rho t}\mathrm{d}t \;\; \text{subject to } \dot{k}(t) = sf(k(t)) - nk(t), \;\; k(0) = k_0,$$

where s is the saving rate, and $k(t)$ is the capital, with n as the parameter of population growth. Find the golden rule optimal saving rate s.

Solution. State Hamiltonian

$$\mathcal{H}(s, k, \lambda) = (1-s)f(k)e^{-\rho t} + \lambda(sf(k) - nk).$$

The FOCs yields

$$\frac{\partial \mathcal{H}}{\partial s} = -f(k)e^{-\rho t} + \lambda f(k) = 0 \tag{4a}$$

$$\frac{\partial \mathcal{H}}{\partial k} = (1-s)f'(k)e^{-\rho t} + \lambda(sf'(k) - n) = -\dot{\lambda} \tag{4b}$$

$$\frac{\partial \mathcal{H}}{\partial \lambda} = sf(k) - nk = \dot{k}. \tag{4c}$$

By (4a),

$$\lambda(t) = e^{-\rho t} \Rightarrow \dot{\lambda}(t) = -\rho e^{-\rho t} = -\rho\lambda.$$

Hence by (4b),

$$(1-s)f'(k)\lambda + \lambda(sf'(k) - n) = \rho\lambda \Rightarrow f'(k^*) = n + \rho,$$

which yields the golden rule capital accumulation. Note that the optimal $k^* = (f')^{-1}(n+\rho)$ is a constant. Hence $\dot{k} = 0$. Therefore the optimal saving rate is

$$s = \frac{nk^*}{f(k^*)}.$$

■

Consider the following continuous time optimal control problem.

$$\max_{x(t), t\geq 0} \int_0^T F(x, y, t)e^{-\rho t} \mathrm{d}t \quad \text{subject to } \dot{y}(t) = g(x, y, t).$$

This is a subclass of the optimal control problem where we explicitly write out the discount factor. We write the Lagrangian as before,

$$\mathcal{L} = \int_0^T \left[F(x, y, t)e^{-\rho t} + \lambda(t)g(x(t), y(t), t) + \dot{\lambda}(t)y(t) \right] \mathrm{d}t - \lambda(t)y(t)\Big|_0^T.$$

Instead of maximizing the end-time outcome, we can consider the present-value optimization by multiplying the whole objective function by $e^{-\rho t}$. That is,

$$\tilde{\mathcal{L}} = \int_0^T \left[F(x, y, t) + \lambda(t)e^{\rho t}g(x(t), y(t), t) + \dot{\lambda}(t)e^{\rho t}y(t) \right] \mathrm{d}t - e^{\rho t}\lambda(t)y(t)\Big|_0^T.$$

Let $\mu(t) = \lambda(t)e^{\rho t}$. Then $\dot{\mu}(t) = \dot{\lambda}(t)e^{\rho t} + \rho\lambda(t)e^{\rho t} = \dot{\lambda}(t)e^{\rho t} + \rho\mu(t)$. Hence we can further rewrite the Lagrangian as

$$\tilde{\mathcal{L}} = \int_0^T \left[F(x,y,t) + \mu(t)g(x(t),y(t),t) + \left(\dot{\mu}(t) - \rho\mu(t)\right)y(t)\right]\mathrm{d}t - \mu(t)y(t)\Big|_0^T.$$

Following the same argument, we can optimize the integrand and find the FOCs,

$$\tilde{\mathcal{I}} = F(x,y,t) + \mu g(x,y,t) + \left(\dot{\mu} - \rho\mu\right)y$$

$$\frac{\partial\tilde{\mathcal{I}}}{\partial x} = F_x(x,y,t) + \mu g_x(x,y,t) = 0$$

$$\frac{\partial\tilde{\mathcal{I}}}{\partial y} = F_y(x,y,t) + \mu g_y(x,y,t) + \dot{\mu} - \rho\mu = 0$$

Again, the solution depends on the derivative of $F + \mu g$. Hence we provide an alternative version of Hamiltonian.

Proposition 2 (Present-value Hamiltonian)

Let the optimal control problem be

$$\max_{x(t)} \int_0^T F(x(t),y(t),t)e^{-\rho t}\mathrm{d}t \quad \text{subject to } \dot{y}(t) = g(x(t),y(t),t),$$

where f is integrable and satisfies necessary regularity conditions. Define the **present-value Hamiltonian** as

$$\mathcal{H}(x,y,\lambda) = F(x,y,t) + \mu g(x,y,t).$$

Then the solution of the optimal control problem satisfies the following first order conditions:

$$\begin{aligned}\frac{\partial\mathcal{H}}{\partial x} &= 0\\ \frac{\partial\mathcal{H}}{\partial y} &= \rho\mu - \dot{\mu}\\ \frac{\partial\mathcal{H}}{\partial \mu} &= g(x,y,t)(=\dot{y}).\end{aligned}$$

We can solve Example 10 again with the present-value Hamiltonian. First state the Hamiltonian and the FOCs:

$$\mathcal{H}(s,k,\mu) = (1-s)f(k) + \mu(sf(k) - nk).$$

$$\frac{\partial \mathcal{H}}{\partial s} = -f(k) + \mu f(k) = 0 \tag{5a}$$

$$\frac{\partial \mathcal{H}}{\partial k} = (1-s)f'(k) + \mu(sf'(k) - n) = \rho\mu - \dot{\mu} \tag{5b}$$

$$\frac{\partial \mathcal{H}}{\partial \mu} = sf(k) - nk = \dot{k}. \tag{5c}$$

Hence $\mu = 1$ and thus $\dot{\mu} = 0$ (from (5a)). Then $f'(k) = n + \rho$ (from (5b)). You might notice that the solution remains the same, but the discount rate is incorporated in the Lagrange multiplier, μ.