

Lecture Note on Production Theory

Kota Saito
California Institute of Technology

September 15, 2026

Contents

1	Introduction: What Is a Firm?	3
1.1	The Black-Box View	3
1.2	Inputs	3
1.3	Time Horizon	3
1.4	Structure of the Lecture	3
2	Single-Input Firm	4
2.1	The Production Function	4
2.2	Productivity Measures	5
2.3	Profit Maximization	6
2.3.1	Geometric Intuition	6
2.4	Cost Function and Supply Curve	7
2.4.1	Time Horizon	7
2.4.2	Types of Costs	8
2.4.3	Marginal Cost	9
2.5	Average Cost	9
2.5.1	The Supply Curve	11
2.6	Producer Surplus	13
2.6.1	Example	13
2.7	Numerical Example: Derivation of Cost Functions from the Production Function . .	14
2.8	Application: Marginal Cost of Electric Power	17
2.8.1	Example: Tohoku Electric Power Company, Japan	17
3	Firm with Multiple Inputs	19
3.1	Returns to Scale	19
3.1.1	Which Returns to Scale Make Sense?	20
3.2	Isoquants and the Marginal Rate of Technical Substitution	21
3.3	Profit Maximization and Cost Minimization	22
3.3.1	The Profit Maximization Problem	22
3.3.2	The Cost Minimization Problem	22
3.4	Properties of the Cost Function and Conditional Factor Demand	24
3.4.1	Revealed Preference in Cost Minimization	28

4	Other Topics	29
4.1	Income Distribution Under CRS	29
4.2	International Productivity and Factor Price Comparisons	29
4.3	Recovering the Production Function from the Profit Function	30
4.3.1	The Legendre–Fenchel Transform	31
4.3.2	Application to Production and Profit	31
4.4	Applicability of Production Theory to Discrete Data	32
5	General Production Sets	35
5.1	Production Sets and Transformation Functions	35
5.2	Properties of Production Sets	36
5.3	Profit Maximization	37
5.4	The Law of Supply	40
5.5	The Weak Axiom of Profit Maximization (WAPM)	41
5.6	Recoverability of the Production Set	41
6	Comparison: Producer Theory and Consumer Theory	42
7	Summary of Key Results	44
A	Duality	44
A.1	Recovering the Technology from the Cost Function	45
A.2	Sufficient Conditions for Cost Functions	45
A.3	Recovering Technology: An Example	46
A.4	Completeness of Restrictions on Factor Demands	46
A.5	Geometry of Duality	46
B	Efficient Production and Welfare	47
C	Aggregation	47
C.1	Necessary and Sufficient Conditions for Profit-Maximizing Behavior	48
C.1.1	Characterization of Profit Functions	48
C.1.2	Characterization of Net Supply Functions	49
C.1.3	Characterization from Discrete Data (WAPM)	50

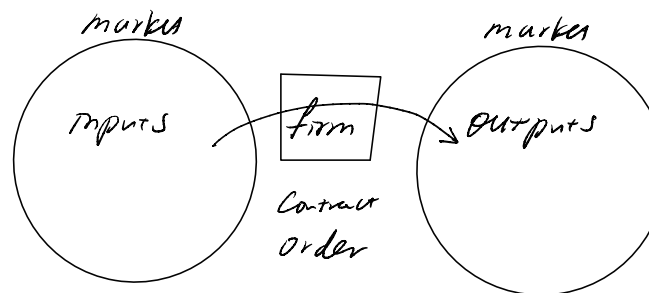
1 Introduction: What Is a Firm?

1.1 The Black-Box View

In microeconomic theory, a *firm* is modeled as an entity that transforms inputs (factors of production) into outputs (goods and services). We deliberately abstract away from the internal organization of the firm—the hierarchy, contracts, incentive structures, and human relationships that govern internal resource allocation. This “black-box” approach allows us to build a minimal conceptual apparatus that is sufficient to analyze *market behavior*.

Methodological note

The internal resource allocation inside a firm is governed by contracts, orders, customs, and human relationships—not by market prices. Game theory and organizational economics study these internal mechanisms. Our focus here is on what the firm does in the market: how it chooses inputs and outputs to maximize profit.



1.2 Inputs

- **Labor (L):** Human resources, typically owned by employees and rented by the firm at wage w .
- **Capital (K):** Machines, buildings, computers, etc., owned by the firm or rented at price r .

1.3 Time Horizon

- **Short run:** At least one input (typically capital) is fixed. The firm can only adjust labor.
- **Long run:** All inputs are variable. The firm can adjust both labor and capital.

1.4 Structure of the Lecture

We proceed in three stages of increasing generality:

1. Firm with a single input (labor)—short-run analysis.
2. Firm with multiple inputs (labor and capital)—long-run analysis.
3. The general production set framework with multiple inputs and multiple outputs.

2 Single-Input Firm

We consider a firm that uses a single input, labor $L \geq 0$, to produce a single output $y \geq 0$. This can be interpreted as short-run analysis in which the capital stock $\bar{K}$ is fixed.

2.1 The Production Function

Definition 2.1 (Production set and production function). The *production set* is

$$\mathcal{Y} = \{(y, L) \in \mathbb{R}_+ \times \mathbb{R}_+ : \text{the firm can produce } y \text{ using labor } L\}.$$

The *production function* $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ describes the boundary of $\mathcal{Y}$:

$$y = f(L) = \max \{y \geq 0 : (y, L) \in \mathcal{Y}\}.$$

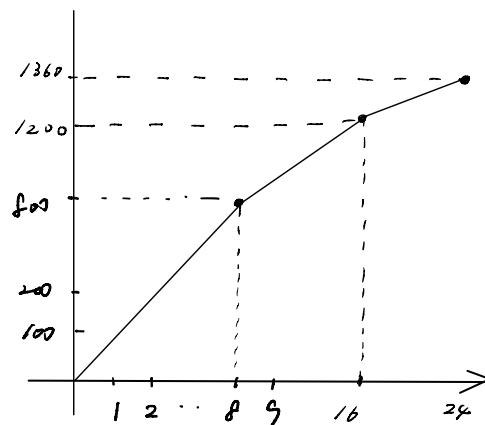
Example 2.1 (Motivating example: CEO with three machines). You are CEO of a small firm making one type of product using labor as the single input.

- You can get \$0.10 by making each product.
- You can hire workers. Each worker can work up to 8 hours and their wage is \$7 per hour.
- You have three machines available for one worker to use:
 - Machine A: 100 products/hour (one worker, up to 8 hours).
 - Machine B: 50 products/hour (one worker, up to 8 hours).
 - Machine C: 20 products/hour (one worker, up to 8 hours).

Q: What is your plan to produce the products?

The firm should assign workers to the most productive machine first. The production function is piecewise linear:

$$f(L) = \begin{cases} 100L & \text{if } 0 \leq L \leq 8, \\ 800 + 50(L - 8) & \text{if } 8 < L \leq 16, \\ 1200 + 20(L - 16) & \text{if } 16 < L \leq 24. \end{cases}$$



2.2 Productivity Measures

Definition 2.2 (Marginal productivity).

$$MP_L(L) = f'(L) = \frac{df}{dL}.$$

Law of diminishing marginal productivity

Under the assumption that labor is the single variable input (capital is fixed), the marginal product of labor is eventually decreasing: $f''(L) < 0$ for L sufficiently large. Economically, as the firm adds workers to a fixed amount of capital, each additional worker contributes less to output. Marginal productivity typically decreases when labor is the single input, or when inputs other than labor (such as machines) cannot be changed in the short period.

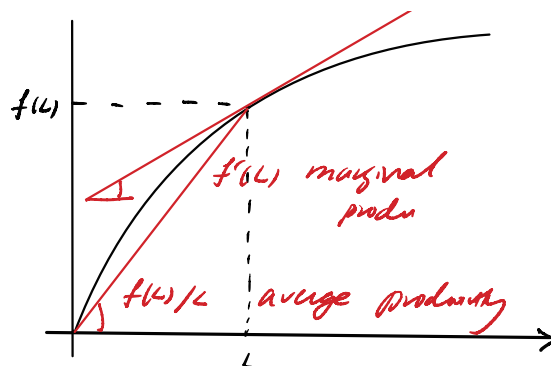


Figure 1: Marginal productivity at L equals to the slope of a tangent line to the production function at L .

2.3 Profit Maximization

Definition 2.3 (Perfect competition). Under *perfect competition*, there are sufficiently many firms and consumers that no single agent can influence market prices. Each firm and consumer is a *price taker*: they treat the output price $p > 0$ and the wage $w > 0$ as exogenously given.

The firm's profit maximization problem is:

$$\max_{L \geq 0} \pi(L) = p f(L) - w L. \quad (1)$$

Proposition 2.1 (First-order condition). If $L^* > 0$ solves (1) and f is differentiable, then

$$p f'(L^*) = w \iff f'(L^*) = \frac{w}{p}. \quad (2)$$

That is, the value of the marginal product equals the wage, or equivalently the marginal product equals the real wage w/p .

Proof. The first-order necessary condition $\pi'(L^*) = 0$ gives $p f'(L^*) - w = 0$. The second-order sufficient condition $\pi''(L^*) = p f''(L^*) < 0$ holds whenever f is strictly concave. $\square$

Remark 2.1. If the value of the marginal input (LHS: $p f'(L)$) exceeds the cost of the marginal input (RHS: w), then the firm can increase profits by using additional input.

Exercise 2.1 (Quiz: Technical innovation). Consider a firm that operates with a production function $y = f(L)$. Thanks to recent technical innovation, the firm can now produce more: the new production function g satisfies $g(L) > f(L)$ for all $L > 0$. The firm is a price taker. At the market price, does this technical innovation necessarily lead to a larger production? Answer yes or no and prove your answer.

2.3.1 Geometric Intuition

Iso-profit lines

Define profit $\pi = py - wL$. Rewriting:

$$y = \underbrace{\frac{w}{p}}_{\text{slope}} L + \underbrace{\frac{\pi}{p}}_{\text{intercept}} \quad (\text{iso-profit line}).$$

This is a family of lines in the (L, y) -plane with slope w/p (the real wage) and vertical intercept π/p . Profit maximization amounts to choosing the highest iso-profit line that still touches the graph of f .

Notice that the optimality condition $f'(L^*) \approx w/p$ is satisfied in the motivating example (Example 2.1). The answer is work 8 hours and use machine A. The iso profits line becomes

$$y = 70 * \text{labor (in hours)} + 10 * \pi.$$

Even though the CEO of the firm does not explicitly solve the model, they behave *as if* doing so. The model is a useful approximation of the decision-making process.

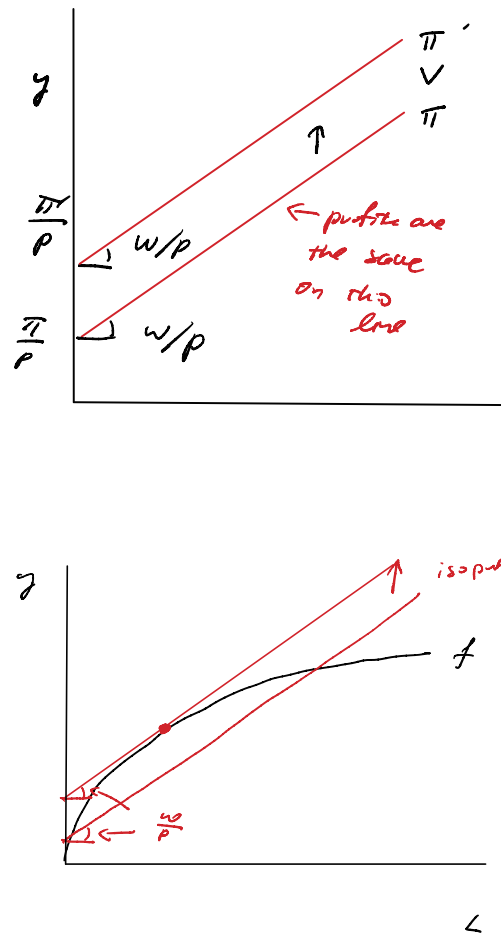


Figure 2: The profit maximization can be understood geometrically as choosing highest iso profit line on the production set given the prices (p, w) .

2.4 Cost Function and Supply Curve

In this section, we study the cost function, which helps us to study the supply curve of the firm.

2.4.1 Time Horizon

As mentioned, when thinking about costs it is important to understand the time horizon:

- **Long period:** You can change all inputs; there is no fixed cost.
- **Short period:** You cannot change all inputs; there is a *fixed cost*—a cost that must be paid in a fixed amount independent of the production level.

Example (pop-up coffee store):

- **Long period:** You are planning to open a one-day pop-up coffee store in a park for Memorial Day. You have not rented a space yet.
- **Short period:** You have already rented a space to open a one-day pop-up store for Memorial Day. You already paid the cost of renting the space and cannot recover it.

2.4.2 Types of Costs

Definition 2.4 (Labor requirement function). The *labor requirement function* $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the inverse of the production function:

$$L(y) = f^{-1}(y) = \min \{L \geq 0 : f(L) \geq y\}.$$

It gives the minimum amount of labor needed to produce y units of output.

Proposition 2.2 (Marginal labor requirement). *If f is strictly increasing and differentiable with $f'(L) > 0$, then by the inverse function theorem:*

$$L'(y) = \frac{1}{f'(L(y))}. \quad (3)$$

Economically, $L'(y)$ is the additional labor needed to produce one more unit of output.

Proof. Differentiating both sides of $f(L(y)) = y$ with respect to y :

$$f'(L(y)) \cdot L'(y) = 1 \quad \implies \quad L'(y) = \frac{1}{f'(L(y))}.$$

□

Proposition 2.3 (Convexity of $L(y)$). *If f is strictly concave ($f'' < 0$), then $L(y)$ is strictly convex.*

Proof. Differentiating (3) with respect to y :

$$L''(y) = \frac{-f''(L(y))}{[f'(L(y))]^3} \cdot L'(y) \cdot f'(L(y)) = \frac{-f''(L(y))}{[f'(L(y))]^3}.$$

Since $f'' < 0$ and $f' > 0$, the numerator is positive and the denominator is positive, so $L''(y) > 0$. □

In the short run, there may be costs that cannot be adjusted:

Definition 2.5 (Cost structure).

- (i) *Sunk cost* $S \geq 0$: Must be paid even if $y = 0$; cannot be recovered. (Example: cost of renting the space in the park.)
- (ii) *Fixed (but not sunk) cost* $F(y) = F \cdot \mathbf{1}_{y>0}$: Must be paid if and only if $y > 0$. (Example: before making coffee on the day, you need to build a tent and clean up the space.)
- (iii) *Variable cost* $wL(y)$: Cost that varies with output, where $L(y) = f^{-1}(y)$. (Example: cost of hiring employees or how many hours you work.)

The *total cost function* is:

$$C(y) = wL(y) + F \cdot \mathbf{1}_{y>0} + S = wL(y) + F \cdot \mathbf{1}_{y>0} + S. \quad (4)$$

It is important to understand how the cost function is related to the production function.

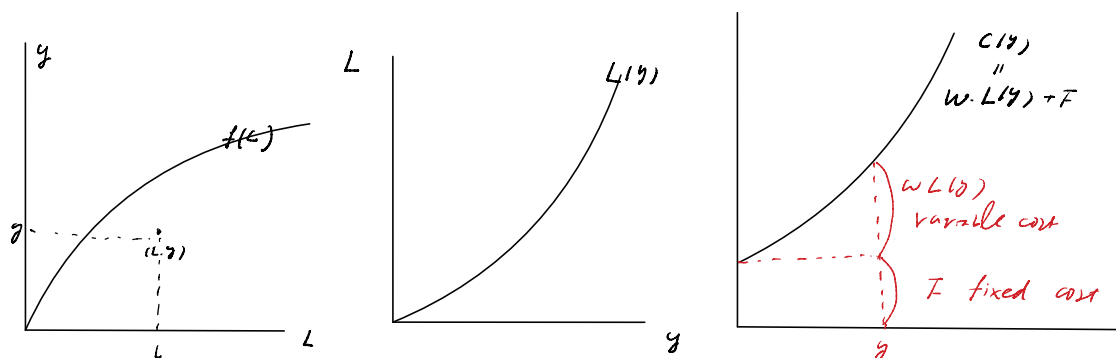


Figure 3: Production function $f(L)$ (left); Labor requirement function $L(y) \equiv f^{-1}(y)$ (center); cost function $C \equiv wL(y) + F$ (right)

2.4.3 Marginal Cost

Definition 2.6 (Marginal Cost).

$$MC(y) = C'(y) = L'(y)w \quad (5)$$

Marginal costs measure an additional amount of cost to produce additional amount of output.

Increasing MC

The strict convexity of $L(y)$ says that each additional unit of output requires *more* additional labor than the last. This is the “quantity” counterpart of the “price” statement that MC is increasing:

$$L''(y) > 0 \iff L'(y) = \frac{1}{f'(L(y))} \text{ is increasing} \iff MC(y) = wL'(y) \text{ is increasing.}$$

Since $C(y) = wL(y) + \text{const}$, the convexity of L is directly inherited by C , which in turn ensures that the supply curve $p = MC(y)$ is upward-sloping as we see later.

2.5 Average Cost

Definition 2.7 (Average Costs). For $y > 0$:

$$AC(y) = \frac{C(y)}{y} = \frac{wL(y) + F + S}{y}, \quad (6)$$

$$AVC(y) = \frac{C(y) - S}{y} = \frac{wL(y) + F}{y}. \quad (7)$$

Proposition 2.4 (Relationship between MC and AC).

$$(i) \quad MC(y) = \frac{w}{f'(L(y))}.$$

$$(ii) \frac{d}{dy}(AC(y)) = \frac{MC(y) - AC(y)}{y}.$$

(iii) AC achieves its minimum at $\hat{y}$ where $MC(\hat{y}) = AC(\hat{y})$. The AC curve is U-shaped.

(iv) Similarly, AVC achieves its minimum at $\tilde{y}$ where $MC(\tilde{y}) = AVC(\tilde{y})$, and $AVC(y) < AC(y)$ for all $y > 0$.

Because of fixed cost, AC is higher than MC at first. Then AC decreases. At some point AC becomes equal to MC . After that point, AC increases because marginal productivity decreases. Thus AC is typically U-shaped, and AC and MC cross at the y that achieves the lowest AC .

Proof.

Part (i): Since $y = f(L)$, we have $dy = f'(L) dL$, so $dL/dy = 1/f'(L)$. Then $MC(y) = w \cdot dL/dy = w/f'(L(y))$.

Part (ii):

$$\frac{d}{dy} AC(y) = \frac{d}{dy} \frac{C(y)}{y} = \frac{C'(y) \cdot y - C(y)}{y^2} = \frac{MC(y) - AC(y)}{y}.$$

Part (iii): By (ii), at any local optimum point $\hat{y} > 0$ of AC , $MC(\hat{y}) = AC(\hat{y})$. It can be shown that AC is convex assuming $f'' < 0$ and other conditions.

We will show $\lim_{y \rightarrow 0^+} AC(y) = +\infty$ and $\lim_{y \rightarrow +\infty} AC(y) = +\infty$ as follows.

For the first limit: For $y > 0$, $AC(y) = \frac{wL(y) + F + S}{y} \geq \frac{F + S}{y}$. Since $F + S > 0$, the right-hand side diverges to $+\infty$ as $y \rightarrow 0^+$. For the second limit: By L'Hôpital's rule:

$$\lim_{y \rightarrow \infty} \frac{L(y)}{y} = \lim_{y \rightarrow \infty} L'(y) = \lim_{y \rightarrow \infty} \frac{1}{f'(L(y))} = +\infty.$$

It follows that $AC(y) \geq wL(y)/y \rightarrow +\infty$.

Part (iv): Can be proved in the same way. □

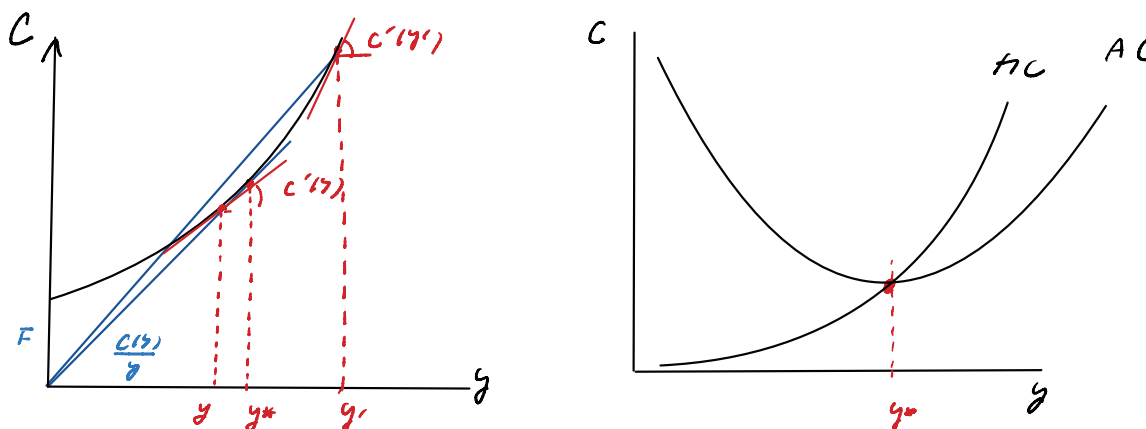


Figure 4: Left figures graphically shows that $AC(y)$ is very high when y is close to zero (because of it is close to $(F + S)/y$) and then decreases until y^* ; at y^* , AC coincides with MC . After the point, AC starts increasing again. This explains the right figure in which AC is U-shaped and MC goes through the lowest point of AC curve.

2.5.1 The Supply Curve

The firm's supply problem, stated in terms of the cost function, is:

$$\max_{y \geq 0} py - C(y). \quad (8)$$

Proposition 2.5 (Equivalence of the two profit maximization formulations (without fixed costs and sunk cost)). *Consider the following two problems:*

$$\max_{L \geq 0} pf(L) - wL, \quad (9)$$

$$\max_{y \geq 0} py - C(y), \quad (10)$$

where $C(y) = wL(y)$ (without fixed costs and sunk cost) and $L(y) = f^{-1}(y)$. These two problems are equivalent:

(i) **Same optimal profit:** $\sup_{L \geq 0} [pf(L) - wL] = \sup_{y \geq 0} [py - C(y)]$.

(ii) **Optimal solutions correspond:** If L^* solves Problem 1, then $y^* = f(L^*)$ solves Problem 2. Conversely, if y^* solves Problem 2, then $L^* = f^{-1}(y^*)$ solves Problem 1.

(iii) **First-order conditions correspond:** The FOC of Problem 1, $pf'(L^*) = w$, and the FOC of Problem 2, $p = MC(y^*)$, are the same condition.

Proof. The key idea is to decompose Problem 1 into two stages: first minimize cost for each output level, then choose the profit-maximizing output.

Step 1: Introduce output as an auxiliary variable. For any $L \geq 0$, let $y = f(L)$. Then optimizing over L is equivalent to first choosing a target output y , then finding the cheapest L that achieves y :

$$\max_{L \geq 0} pf(L) - wL = \max_{y \geq 0} \left(\max_{L: f(L)=y} py - wL \right). \quad (11)$$

The equality holds because every L corresponds to some $y = f(L)$, and every $y \geq 0$ in the range of f has at least one L with $f(L) = y$.

Step 2: The inner problem defines the cost function. In the inner maximization, y is fixed, so py is a constant with respect to L . Therefore:

$$\max_{L: f(L)=y} py - wL = py - \min_{L: f(L)=y} wL = py - wf^{-1}(y) = py - C(y),$$

where the last equality uses $C(y) = wL(y)$.

Step 3: Substituting back. Combining Steps 1 and 2:

$$\max_{L \geq 0} pf(L) - wL = \max_{y \geq 0} [py - C(y)].$$

This proves (i) and imply (ii).

Proof of (iii): The FOC of Problem 1 is:

$$pf'(L^*) = w \quad \Longleftrightarrow \quad p = \frac{w}{f'(L^*)}.$$

The FOC of Problem 2 is:

$$p = C'(y^*) = MC(y^*).$$

By Proposition 2.4(i), $MC(y) = w/f'(L(y))$. Since $L^* = L(y^*)$ by part (ii), the two conditions are identical:

$$p = \frac{w}{f'(L^*)} = \frac{w}{f'(L(y^*))} = MC(y^*). \quad \square$$

Theorem 2.1 (Supply curve of the competitive firm). Define the close-down price $p^* = \min_y AVC(y)$ and the break-even price $p^{**} = \min_y AC(y)$.

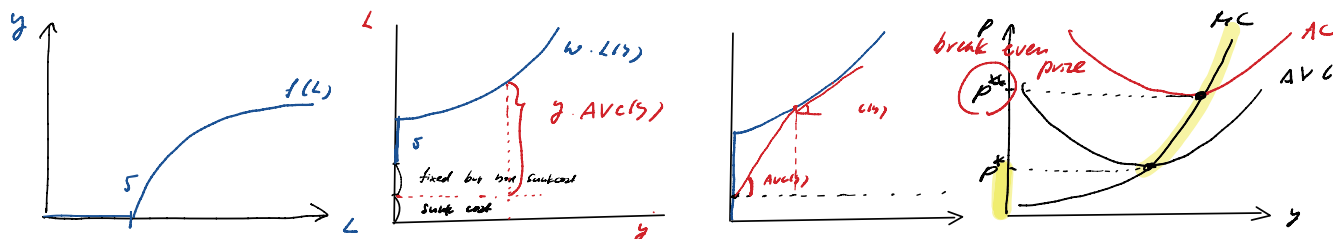
- (i) If $p > p^*$, the firm produces $y > 0$ satisfying $p = MC(y)$, i.e., $S(p) = MC^{-1}(p)$.
- (ii) If $p < p^*$, the firm shuts down: $S(p) = 0$.
- (iii) For $p^* < p < p^{**}$, the firm produces but earns negative profit (it still covers variable costs).
- (iv) For $p > p^{**}$, the firm earns strictly positive profit.

Proof. The profit from producing $y > 0$ is $py - wL(y) - F - S = (p - AVC(y))y - S$. The profit from shutting down ($y = 0$) is $-S$. The firm produces if and only if

$$(p - AVC(y))y > 0 \quad \text{for some } y > 0.$$

If $p < p^* = \min_y AVC(y)$, then $p < AVC(y)$ for all y , so $(p - AVC(y))y < 0$ for all $y > 0$. Hence the firm shuts down. If $p > p^*$, there exists y with $p > AVC(y)$, so production is profitable relative to variable costs. The FOC $p = MC(y)$ determines the optimal output.

□



In the picture above, the production function is not concave because we are assuming that the firm needs to operate fixed hours to produce a product as long as the production level is positive.¹

Remark 2.2. In the pop-up coffee store example:

- p^* is the price below which it is better not to build the tent and clean up the space (i.e., not to produce at all).
- p^{**} is the price below which your net profit (including the cost of renting the space) is negative. However, once you have rented the space, if the price is above p^* (even if below p^{**}), it is better to build the tent and make coffee.

¹All results hold even without this assumption.

2.6 Producer Surplus

Definition 2.8 (Producer surplus). The *producer surplus* is the profit before subtracting sunk costs:

$$PS = p y - w L(y) - F = y(p - AVC(y)) = \underbrace{py}_{\text{Sale}} - \underbrace{y AVC(y)}_{\text{Variable costs}}.$$

Graphically, it is the area between the supply curve and the price line:

$$PS = p y^* - \int_0^{y^*} MC(y) dy,$$

where $y^* = S(p)$.

Remark 2.3. Equivalently, using integration by parts on the supply curve $S(\cdot)$:

$$PS(p) = \int_{p^*}^p S(\tilde{p}) d\tilde{p}.$$

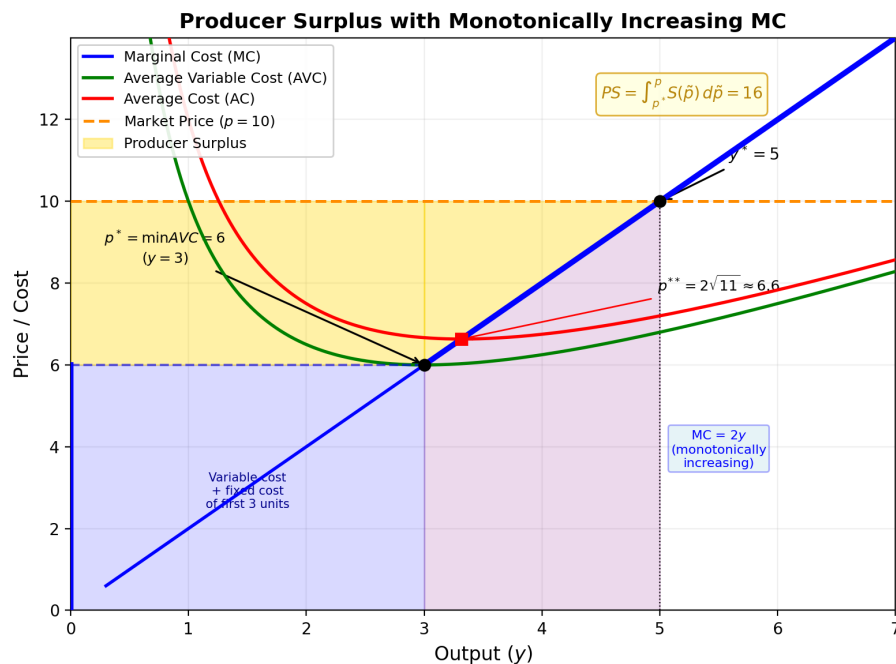
This mirrors the consumer surplus formula in demand theory.

2.6.1 Example

In the example below, the MC and AVC cross at $y = 3$ and $p = 2$ (i.e., $MC(3) = AVC(3) = 2$). Thus, the supply curve is given by

$$S(p) = \begin{cases} 0 & \text{if } p \leq 2, \\ MC^{-1}(p) & \text{if } p > 2. \end{cases}$$

Suppose that the price is $p = 10$ and the output is $y^* = 5$. The producer surplus is the area between the supply curve and the price line (the yellow area in the figure below).



To see this, notice that:

- (Sales of selling 5 outputs with price 10) = (Rectangle with height 10 and width 5).
- (Variable costs of producing 5) = (Blue area) + (Purple area):
 - Blue area: Variable costs of producing 3 outputs (i.e., $AVC(3) \times 3$).
 - Purple area: The sum of marginal costs of producing the 4th output (from the 3rd) and the 5th output (from the 4th).
- Thus, (Producer surplus) = (Rectangle) – (Blue area) – (Purple area) = (Yellow area).

2.7 Numerical Example: Derivation of Cost Functions from the Production Function

Given a production function f , a wage w , a fixed cost F , and a sunk cost S , we derive all cost functions. The logical chain is:

$$f(L) \xrightarrow{\text{invert}} L(y) \xrightarrow{\times w} V(y) \xrightarrow{+F, S} C(y) \xrightarrow{\text{differentiate, divide}} MC(y), AVC(y), AC(y).$$

Step 1: The Production Function

The firm requires $L_0 = 4$ hours of preparation before it can produce any output. After preparation, output is concave in additional labor. Formally:

$$f(L) = \begin{cases} 0 & \text{if } L \leq 4, \\ \sqrt{L-4} & \text{if } L > 4. \end{cases}$$

This is a concave function on $(4, \infty)$ since $f''(L) = -\frac{1}{4}(L-4)^{-3/2} < 0$ for $L > 4$.

Remark 2.4. The preparation time $L_0 = 4$ is a *labor startup cost*: the worker must be paid for 4 hours before any output is produced. Economically, this could represent setting up equipment, organizing the workspace, or training.

Step 2: The Labor Requirement Function

Inverting $y = \sqrt{L-4}$ gives the labor required to produce $y > 0$ units of output:

$$L(y) = 4 + y^2.$$

Step 3: The Variable Cost Function

The variable cost is the wage bill. Normalizing $w = 1$:

$$w \cdot L(y) = 4 + y^2 \quad (y > 0), \quad V(0) = 0.$$

Note that $V(y)$ is *discontinuous* at $y = 0$: the moment the firm produces any positive amount, it must pay the preparation cost $wL_0 = 4$.

Step 4: Adding Fixed and Sunk Costs

In addition to the labor cost, the firm faces:

- **Fixed (but not sunk) cost** $F = 5$: paid if and only if $y > 0$ (e.g., setting up a tent, cleaning the space).
- **Sunk cost** $S = 2$: paid regardless of whether the firm produces (e.g., cost of renting the space, already paid).

The **total cost function** is:

$$C(y) = \underbrace{(4 + y^2)}_{\text{variable cost } V(y)} + \underbrace{5 \cdot \mathbf{1}_{y>0}}_{\text{fixed cost } F} + \underbrace{2}_{\text{sunk cost } S} = y^2 + 9 \cdot \mathbf{1}_{y>0} + 2.$$

For $y > 0$:

$$C(y) = y^2 + 11.$$

Step 5: Deriving Marginal Cost

The marginal cost is the derivative of total cost with respect to output (for $y > 0$):

$$MC(y) = C'(y) = V'(y) = \frac{dV}{dy} = 2y.$$

Alternatively, from the production function directly:

$$MC(y) = \frac{w}{f'(L(y))} = \frac{1}{\frac{1}{2}(L(y) - 4)^{-1/2}} = 2\sqrt{L(y) - 4} = 2y.$$

Since f is concave (so $f' > 0$ is decreasing), MC is **monotonically increasing** in y . This is in contrast to the S-shaped production function case, where MC is U-shaped.

Step 6: Average Cost Curves

Average variable cost (includes V and F , excludes S):

$$AVC(y) = \frac{V(y) + F}{y} = \frac{4 + y^2 + 5}{y} = \frac{9}{y} + y.$$

Average cost (includes everything):

$$AC(y) = \frac{V(y) + F + S}{y} = \frac{4 + y^2 + 5 + 2}{y} = \frac{11}{y} + y.$$

Step 6a: Why AVC Is U-shaped Even Though MC Is Increasing

AVC is U-shaped because of the preparation cost and fixed cost being amortized over output:

$$AVC'(y) = -\frac{9}{y^2} + 1.$$

Setting $AVC'(y) = 0$ gives $y^2 = 9$, so $y = 3$. The minimum average variable cost is:

$$AVC(3) = \frac{9}{3} + 3 = 6 = MC(3).$$

For small y , the term $9/y$ dominates (spreading startup and fixed costs over few units), pulling AVC above MC . For large y , the y term dominates, and AVC approaches MC from below. The key insight: *the U-shape of AVC comes entirely from amortizing the startup cost $wL_0 + F$, not from any non-monotonicity in MC .*

Step 7: Key Prices and the Supply Curve

Close-down price (minimum of AVC):

$$p^* = \min_y AVC(y) = AVC(3) = 6.$$

Break-even price (minimum of AC):

$$AC'(y) = -\frac{11}{y^2} + 1 = 0 \implies y = \sqrt{11},$$

$$p^{**} = AC(\sqrt{11}) = \frac{11}{\sqrt{11}} + \sqrt{11} = 2\sqrt{11} \approx 6.63.$$

Supply curve:

$$S(p) = \begin{cases} 0 & \text{if } p < 6, \\ MC^{-1}(p) = \frac{p}{2} & \text{if } p \geq 6. \end{cases}$$

Note: the supply curve *jumps* from $y = 0$ to $y = 3$ at $p = p^* = 6$.

Step 8: Producer Surplus at $p = 10$

At $p = 10$, the firm produces $y^* = MC^{-1}(10) = 5$. The producer surplus is:

$$\begin{aligned} PS &= py^* - V(y^*) - F \\ &= 10 \cdot 5 - (4 + 25) - 5 = 50 - 29 - 5 = 16. \end{aligned}$$

Equivalently, using the area to the left of the supply curve:

$$PS = \int_{p^*}^p S(\tilde{p}) d\tilde{p} = \int_6^{10} \frac{\tilde{p}}{2} d\tilde{p} = \left[\frac{\tilde{p}^2}{4} \right]_6^{10} = 25 - 9 = 16.$$

Decomposing into the rectangle and curved parts:

$$PS = \underbrace{3 \times (10 - 6)}_{= 12 \text{ (rectangle)}} + \underbrace{\int_3^5 (10 - 2y) dy}_{= 4 \text{ (triangle)}} = 12 + 4 = 16.$$

In fact, since $MC = 2y$ is linear, the “curved part” is actually a triangle with base $5 - 3 = 2$ and height $10 - 6 = 4$, giving area $\frac{1}{2} \cdot 2 \cdot 4 = 4$.

2.8 Application: Marginal Cost of Electric Power

2.8.1 Example: Tohoku Electric Power Company, Japan

- Fixed cost: expenditure on plant and equipment, labor cost, etc.
- Variable cost:
 - Nuclear power: almost zero (once the critical state is achieved).
 - Hydro power: almost zero (all you need is to open the gate).
 - Thermal power: the cost of fuel (from the company's public information, we know the fuel cost).

Table 1: Cost structure of Tohoku Electric Power Company

Type	Plant	Capacity (MW)	MC (yen/kWh)	Utilization (%)
Nuclear	—	2077	0	100
Hydro	—	2896	0	100
Coal	Haramachi	1896	1.6	81.4
Coal	Noshiro	1134	1.7	77.9
Coal	Sendai	476	1.8	69.5
LNG	East Niigata	3717	5.1	63.3
LNG	Shin Sendai	578	6.3	45.6
LNG	Niigata	477	6.9	34.5
Oil	Shin Sendai	328	7.1	19.3
Oil	Hachinohe	457	7.9	19.5
Oil	Akita	1495	8.3	10.9

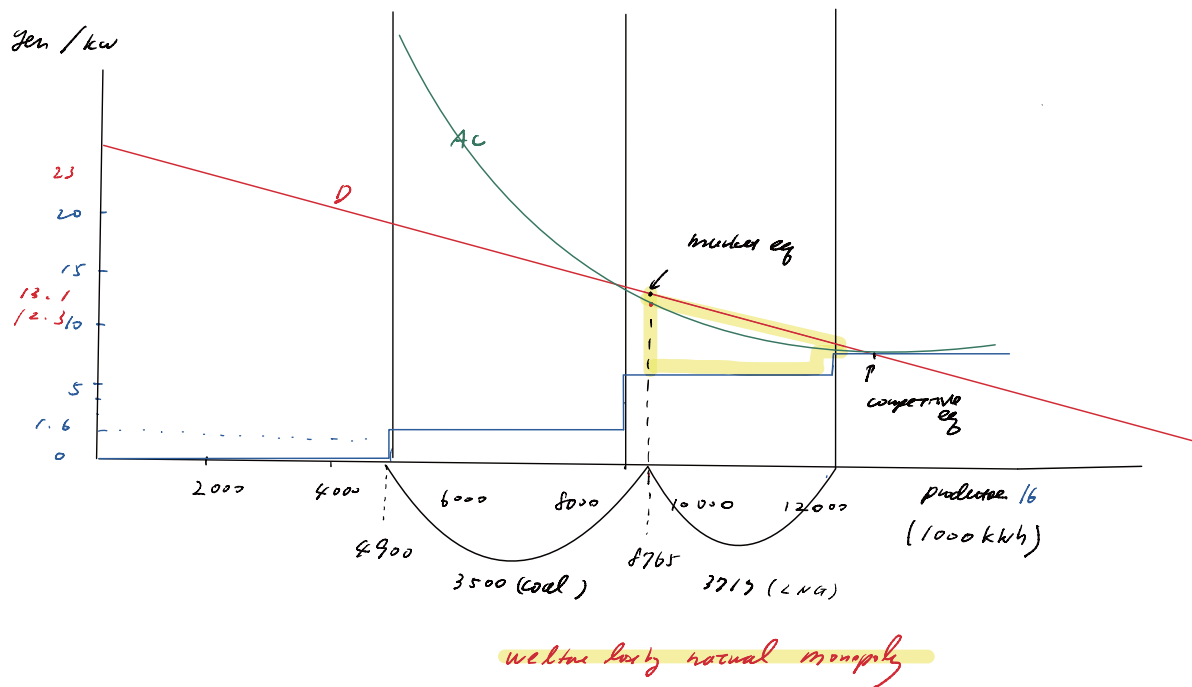
- Notice that the utilization rate is higher for plants with lower marginal cost. (Companies use more efficient power plants with smaller marginal costs first.)
- From the table above, we can write the supply curve.
- They also report that in 2001, the average cost of producing 1 kWh is 12.3 yen and the average production per hour is 8765×1000 kWh. That is, $AC(8765) = 12.3$.
- From this information, we can calculate the fixed cost. Recall that

$$AC(y) = \frac{\int_0^y MC(x) dx + F + S}{y},$$

where F is a fixed but not sunk cost and S is a sunk cost.

- From the figure, $\int_0^{8765} MC(x) dx \approx 7500$. Thus

$$12.3 = AC(8765) = \frac{7500 + F + S}{8765}.$$



- Hence $F + S \approx 100300$.
- From this information, we can draw the AC curve.
- From the figure, we obtain the following insight:

Key Observation

- The firm will operate when the price exceeds AVC . We cannot obtain AVC from the available data, but we know $AC > AVC$; the supply curve is a part of the MC curve.
- The actual market price in that year was 13.1 yen (per 1 kWh) and the average production per hour was 8765×1000 kWh.
- Notice that the production quantity 8765 is lower than the level on the supply curve (MC curve), and the price is higher than the level on the supply curve (MC curve).
- This is because the electric power company is *not* a price taker—it has market power.

3 Firm with Multiple Inputs

We now consider a firm with two inputs: labor L and capital K , with a production function

$$y = F(L, K),$$

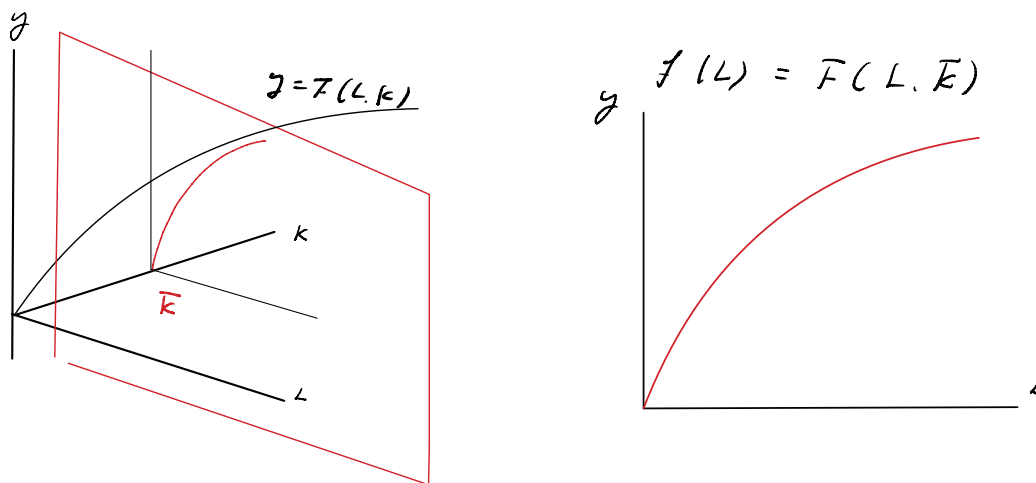
where $F : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is twice continuously differentiable with $F_L > 0$ and $F_K > 0$.

- **Labor** L : can be adjusted in the short period. For example, consider a moving company.
- **Capital** K : fixed production equipment such as trucks and parking places that cannot be adjusted in the short period. Given a fixed capital level $\bar{K}$, the short-period production function (studied in the previous section) is

$$f(L) = F(L, \bar{K}).$$

In the following, we study two new topics that arise with multiple inputs:

- Returns to scale
- Substitution among inputs / technical marginal rate of substitution



3.1 Returns to Scale

We ask the following question: if we increase *all* inputs by a factor $t > 0$, how much does production increase? Note that this is different from the question of how much production increases if we increase only one input, say labor.

Definition 3.1 (Returns to scale). For all $(L, K) \in \mathbb{R}_{++}^2$ and $t > 1$:

- Constant returns to scale (CRS)*: $F(tL, tK) = tF(L, K)$ (F is Homog. of degree 1).
- Increasing returns to scale (IRS)*: $F(tL, tK) > tF(L, K)$.

(iii) *Decreasing returns to scale (DRS)*: $F(tL, tK) < tF(L, K)$.

Example 3.1 (Cobb–Douglas). $F(L, K) = L^\alpha K^\beta$ with $\alpha, \beta > 0$. Then $F(tL, tK) = t^{\alpha+\beta}F(L, K)$. So the technology exhibits CRS if $\alpha + \beta = 1$, IRS if $\alpha + \beta > 1$, and DRS if $\alpha + \beta < 1$.

Example 3.2 (Leontief (fixed proportions)). Let $a, b > 0$. The *Leontief production function* is:

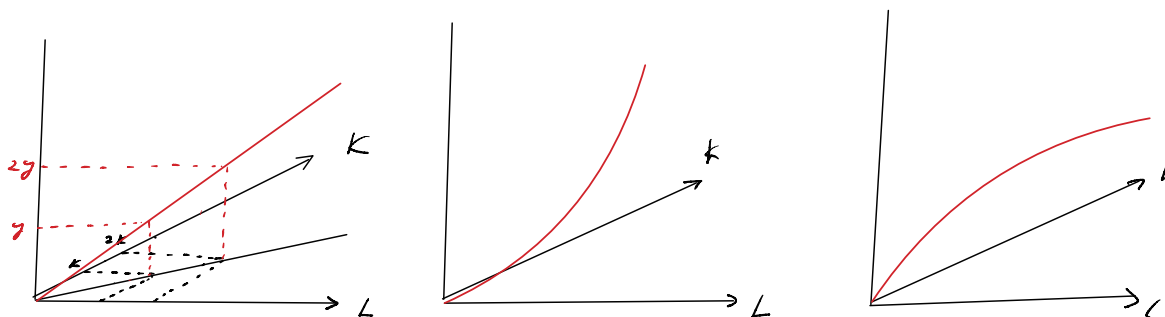
$$F(L, K) = \min(aL, bK).$$

This technology exhibits CRS. The isoquants are L-shaped: to produce y units of output, the firm needs at least $L = y/a$ and $K = y/b$. Any excess of one input beyond this ratio is wasted. This models production processes with rigid input proportions, such as one truck needing exactly one driver.

Example 3.3 (Constant elasticity of substitution (CES)). Let $\alpha > 0$, $\beta > 0$, and $\rho \neq 0$ with $\rho < 1$. The *CES production function* is:

$$F(L, K) = (\alpha L^\rho + \beta K^\rho)^{1/\rho}.$$

Special cases: as $\rho \rightarrow 0$, the CES converges to the Cobb–Douglas $L^{\alpha/(\alpha+\beta)}K^{\beta/(\alpha+\beta)}$; as $\rho \rightarrow -\infty$, it converges to the Leontief $\min\{L, K\}$; and $\rho = 1$ gives the linear (perfect substitutes) technology $\alpha L + \beta K$.



3.1.1 Which Returns to Scale Make Sense?

A few observations:

- **Increasing returns to scale and perfect competition are incompatible.** Under IRS, only the biggest company will win and the market will be monopolized. So we will not consider this case. Thus, in our analysis of competitive markets, we restrict to CRS or DRS.
- **In a very long period, constant returns to scale may make sense.** In a very long period, the firm can adjust all inputs. For example, if the firm makes two plants with exactly the same labor and equipment, then production will be doubled.
- **In a relatively long period, decreasing returns to scale may make sense.** This is because the firm can adjust labor but not all capital. For example, even if you make two plants with exactly the same labor and equipment, production may not double if the ability of the CEO is limited.²

²In the very long run, you can hire a new CEO.

When do different returns apply?

- **Very long run** (all inputs adjustable): CRS is natural—replicating all inputs exactly should replicate output.
- **Relatively long run** (some inputs fixed, e.g., CEO talent): DRS is natural—doubling labor and machines may not double output if managerial capacity is a binding constraint.
- **Short run** (capital fixed): The production function $f(L) = F(L, \bar{K})$ typically exhibits strict concavity in L . Thus, DRS.

Proposition 3.1 (Concavity of the production function). *Suppose F is twice continuously differentiable. If F exhibits DRS (or CRS) and each cross-partial $F_{LK} \geq 0$, and each own-partial second derivative satisfies $F_{LL} < 0$ and $F_{KK} < 0$, then F is concave on $\mathbb{R}_+^2$.*

Proof. The Hessian matrix of F is

$$H_F = \begin{pmatrix} F_{LL} & F_{LK} \\ F_{KL} & F_{KK} \end{pmatrix}.$$

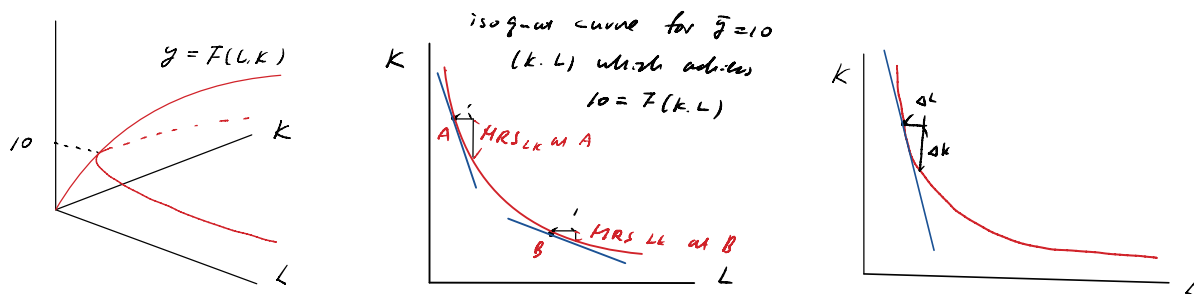
F is concave if and only if H_F is negative semidefinite, which requires $F_{LL} \leq 0$, $F_{KK} \leq 0$, and $\det(H_F) = F_{LL}F_{KK} - (F_{LK})^2 \geq 0$. The conditions stated ensure this. $\square$

3.2 Isoquants and the Marginal Rate of Technical Substitution

Definition 3.2 (Isoquant). For a given output level $\bar{y}$, the *isoquant* is

$$\mathcal{Q}(\bar{y}) = \{(L, K) \in \mathbb{R}_+^2 : F(L, K) = \bar{y}\}.$$

From the curve, we can see how much capital the firm can save by increasing one unit of labor.



Definition 3.3 (Marginal rate of technical substitution). The *marginal rate of technical substitution of labor for capital* is

$$MRTS_{LK}(L, K) = -\frac{dK}{dL} \Big|_{F=\bar{y}} = \frac{\partial F / \partial L}{\partial F / \partial K} = \frac{F_L}{F_K}.$$

It measures how much capital the firm can save by increasing one unit of labor, holding output constant.

Derivation. Totally differentiating $F(L, K) = \bar{y}$:

$$F_L dL + F_K dK = 0 \implies \frac{dK}{dL} = -\frac{F_L}{F_K}.$$

□

Diminishing MRTS

Law of decreasing marginal rate of substitution: $MRTS_{LK}$ decreases as L increases. If the company has enough drivers, adding one more driver does not help much. If F is strictly quasiconcave, then $MRTS_{LK}$ decreases along the isoquant as L increases (and K decreases). The isoquants are convex to the origin.

3.3 Profit Maximization and Cost Minimization

3.3.1 The Profit Maximization Problem

$$\max_{L, K \geq 0} pF(L, K) - wL - rK, \quad (12)$$

where $w > 0$ is the wage and $r > 0$ is the *rental price* of capital (i.e., the price of using capital for one period). For example, if the capital is a truck, then r is the price of the truck divided by the number of days the truck can be used.

Proposition 3.2 (FOC for profit maximization). *At an interior optimum $(L^*, K^*) \gg 0$:*

$$pF_L(L^*, K^*) = w, \quad (13)$$

$$pF_K(L^*, K^*) = r, \quad (14)$$

which together imply

$$MRTS_{LK} = \frac{F_L}{F_K} = \frac{w}{r}. \quad (15)$$

The FOC is sufficient for an interior solution if F is a concave function.

Proof. The gradient of the objective must vanish at an interior maximum: $\nabla_{L,K}[pF(L, K) - wL - rK] = (pF_L - w, pF_K - r) = (0, 0)$. □

Remark 3.1. Conditions (13)–(14) say: the value of the marginal product of each input equals its price. Condition (15) says: the technical rate of substitution equals the economic rate of substitution (the factor price ratio).

3.3.2 The Cost Minimization Problem

Another way of obtaining the same condition is through cost minimization. Let C be a fixed level of cost. Then the *iso-cost line* can be written as

$$K = -\frac{w}{r}L + \frac{C}{r}.$$

Given the isoquant curve with fixed production $\bar{y}$, a cost-minimizing firm will choose the pair (L, K) at which the iso-cost line is tangent to the isoquant curve.

Definition 3.4 (Cost minimization problem (CMP)). For a given output level $y \geq 0$, the CMP is:

$$c(w, r, y) = \min_{L, K \geq 0} wL + rK \quad \text{s.t.} \quad F(L, K) \geq y. \quad (16)$$

The optimized value $c(w, r, y)$ is the *cost function*. The optimizing input vector $\mathbf{z}(w, r, y) = (L(w, r, y), K(w, r, y))$ is the *conditional factor demand*.

Remark 3.2. Conditionally factor demand can be interpreted as a demand for consumer theory

Remark 3.3. In the short run, for the single input case, given $\bar{K}$

$$c(w, y) = \min_{L \geq 0} wL + r\bar{K} \quad \text{s.t.} \quad F(L, \bar{K}) \geq q = w f^{-1}(y) + r\bar{K} = wL(y) + r\bar{K},$$

where $f(L) = F(L, \bar{K})$. Notice that this formula coincides with the definition of the cost function for single input case.

Proposition 3.3 (FOC for cost minimization). If $\mathbf{z}^* = (L^*, K^*) \gg 0$ solves the CMP with f concave, then for some $\lambda \geq 0$:

$$w = \lambda F_L(L^*, K^*), \quad r = \lambda F_K(L^*, K^*), \quad (17)$$

so that

$$\frac{F_L(L^*, K^*)}{F_K(L^*, K^*)} = \frac{w}{r} = MRTS_{LK}.$$

The multiplier $\lambda = \partial c(w, r, y) / \partial q$ is the marginal cost of production.

Proof. Form the Lagrangian $\mathcal{L} = wL + rK - \lambda(F(L, K) - y)$. The FOCs are $w = \lambda F_L$ and $r = \lambda F_K$. The interpretation of λ follows from the envelope theorem. $\square$

In the following, I might have wrote q instead of y .

Iso-cost lines

An iso-cost line for expenditure C is $K = -\frac{w}{r}L + \frac{C}{r}$, a line with slope $-w/r$. Cost minimization picks the lowest iso-cost line tangent to the isoquant $F(L, K) = y$, yielding $MRTS_{LK} = w/r$.

Example 3.4 (Cost minimization for standard technologies).

- (i) **Cobb–Douglas** $F = L^\alpha K^\beta$, $\alpha + \beta = 1$. The FOC $\frac{\alpha K}{\beta L} = \frac{w}{r}$ gives $K = \frac{w\beta}{r\alpha}L$. Substituting into $L^\alpha K^\beta = y$:

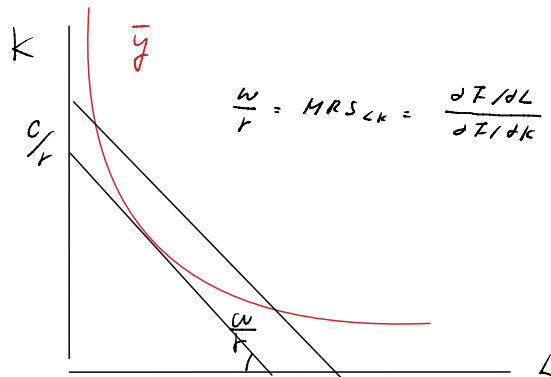
$$c(w, r, y) = y \cdot w^\alpha r^\beta \cdot \left(\frac{\alpha}{\beta}\right)^{-\beta} \cdot \left(1 + \frac{\beta}{\alpha}\right).$$

Under CRS, c is linear in y (constant average cost).

- (ii) **Leontief** $F = \min(aL, bK)$. At the optimum, $aL = bK = y$, so $L = y/a$, $K = y/b$:

$$c(w, r, y) = \left(\frac{w}{a} + \frac{r}{b}\right)y.$$

The cost function is linear in w and r (no substitution possible).



(iii) **Linear** $F = aL + bK$. The firm uses only the cheaper input:

$$c(w, r, y) = \min \left\{ \frac{w}{a}, \frac{r}{b} \right\} y.$$

The cost function is concave in (w, r) —indeed, it is the minimum of two linear functions, which is concave.

Remark 3.4 (Duality of curvature). Notice the pattern: the Leontief production function has *L-shaped* (sharply curved) isoquants but a *linear* cost function; the linear production function has *flat* isoquants but a *kinked* (sharply curved) cost function. This inverse relationship between isoquant curvature and isocost curvature is a manifestation of duality (see Section A).

3.4 Properties of the Cost Function and Conditional Factor Demand

The following results give a characterization of cost function given the production. The result is important because sometimes the observable data is the cost function and factor demand.

Proposition 3.4 (Properties of $c(\cdot)$ and $\mathbf{z}(\cdot)$). *Suppose F is concave and the production set Y is closed and satisfies free disposal. Then:*

- (i) $c(\cdot)$ is Homog. of degree one in (w, r) and Increasing in y .
- (ii) $c(\cdot)$ is a concave function of (w, r) .
- (iii) If $\{(L, K) \geq 0 : F(L, K) \geq y\}$ is a strictly convex set, then $\mathbf{z}(w, r, y)$ is single-valued.
- (iv) (Shephard's lemma) If $\mathbf{z}(\bar{w}, \bar{r}, y)$ is a singleton, then

$$\nabla_{(w,r)} c(\bar{w}, \bar{r}, y) = \mathbf{z}(\bar{w}, \bar{r}, y).$$

- (v) The matrix $D_{(w,r)}^2 c(\bar{w}, \bar{r}, y) = D_{(w,r)} \mathbf{z}(\bar{w}, \bar{r}, y)$ is symmetric and negative semidefinite, with $D_{(w,r)}^2 c \cdot (\bar{w}, \bar{r})^\top = 0$.
- (vi) If F is Homog. of degree one (CRS), then $c(\cdot)$ and $\mathbf{z}(\cdot)$ are Homog. of degree one in y .
- (vii) If F is concave, then $c(\cdot)$ is convex in y ; in particular, marginal costs are Increasing.

Proof. Throughout, the cost minimization problem is

$$c(w, r, y) = \min_{L, K \geq 0} wL + rK \quad \text{s.t.} \quad F(L, K) \geq y,$$

and $\mathbf{z}(w, r, y) = \arg \min\{wL + rK : F(L, K) \geq y, L, K \geq 0\}$ is the set of minimizers.

Part (i): Homogeneity of degree one in (w, r) and monotonicity in y .

Homogeneity. Let $t > 0$. The constraint set $V(y) = \{(L, K) \geq 0 : F(L, K) \geq y\}$ does not depend on factor prices. Therefore:

$$\begin{aligned} c(tw, tr, y) &= \min_{(L, K) \in V(y)} (tw)L + (tr)K \\ &= t \min_{(L, K) \in V(y)} wL + rK \\ &= t \cdot c(w, r, y). \end{aligned}$$

Moreover, the set of minimizers is unchanged: $\mathbf{z}(tw, tr, y) = \mathbf{z}(w, r, y)$.

Monotonicity in y . If $y' > y$, then $V(y') = \{(L, K) : F(L, K) \geq y'\} \subseteq \{(L, K) : F(L, K) \geq y\} = V(y)$. Minimizing over a smaller set yields a weakly larger value:

$$c(w, r, y') = \min_{(L, K) \in V(y')} wL + rK \geq \min_{(L, K) \in V(y)} wL + rK = c(w, r, y).$$

Part (ii): Concavity in (w, r) .

Let (w^0, r^0) and (w^1, r^1) be two factor price vectors, and let $\alpha \in [0, 1]$. Define $(w^\alpha, r^\alpha) = \alpha(w^0, r^0) + (1 - \alpha)(w^1, r^1)$. Let $(L^\alpha, K^\alpha) \in \mathbf{z}(w^\alpha, r^\alpha, y)$ be a cost-minimizing bundle at the convex combination prices. Then:

$$\begin{aligned} c(w^\alpha, r^\alpha, y) &= w^\alpha L^\alpha + r^\alpha K^\alpha \\ &= \alpha(w^0 L^\alpha + r^0 K^\alpha) + (1 - \alpha)(w^1 L^\alpha + r^1 K^\alpha). \end{aligned}$$

Since $(L^\alpha, K^\alpha) \in V(y)$ is feasible but not necessarily cost-minimizing at prices (w^0, r^0) or (w^1, r^1) :

$$w^0 L^\alpha + r^0 K^\alpha \geq c(w^0, r^0, y) \quad \text{and} \quad w^1 L^\alpha + r^1 K^\alpha \geq c(w^1, r^1, y).$$

Substituting:

$$c(w^\alpha, r^\alpha, y) \geq \alpha c(w^0, r^0, y) + (1 - \alpha) c(w^1, r^1, y). \quad \square$$

Part (iii): Strict convexity of $V(y)$ implies $\mathbf{z}$ is single-valued.

Suppose (L^a, K^a) and (L^b, K^b) are both in $\mathbf{z}(w, r, y)$, i.e., both achieve the minimum cost $c(w, r, y)$:

$$wL^a + rK^a = wL^b + rK^b = c(w, r, y).$$

Both bundles lie in $V(y)$. Since $V(y)$ is strictly convex, for any $\alpha \in (0, 1)$, the strict convex combination $(L^\alpha, K^\alpha) = \alpha(L^a, K^a) + (1 - \alpha)(L^b, K^b)$ satisfies $F(L^\alpha, K^\alpha) > y$ (strict inequality, by strict convexity of the upper contour set). By continuity of F , we can reduce (L^α, K^α) slightly to some (L', K') with $F(L', K') \geq y$ and

$$wL' + rK' < wL^\alpha + rK^\alpha = \alpha c(w, r, y) + (1 - \alpha) c(w, r, y) = c(w, r, y).$$

This contradicts the definition of $c(w, r, y)$ as the minimum cost. Therefore $(L^a, K^a) = (L^b, K^b)$.

Part (iv): Shephard's lemma.

We give two proofs.

Proof via the envelope theorem. Form the Lagrangian $\mathcal{L} = wL + rK - \lambda(F(L, K) - y)$. At the optimum (L^*, K^*, λ^*) , the envelope theorem gives:

$$\frac{\partial c}{\partial w} = \frac{\partial \mathcal{L}}{\partial w} \Big|_{(L^*, K^*, \lambda^*)} = L^*, \quad \frac{\partial c}{\partial r} = \frac{\partial \mathcal{L}}{\partial r} \Big|_{(L^*, K^*, \lambda^*)} = K^*.$$

That is, $\nabla_{(w,r)} c(\bar{w}, \bar{r}, y) = (L^*, K^*) = \mathbf{z}(\bar{w}, \bar{r}, y)$.

Proof from first principles (not assuming differentiability of the constraint). Fix output y and a cost-minimizing bundle $(L^*, K^*) = \mathbf{z}(\bar{w}, \bar{r}, y)$. For any factor prices (w, r) , the bundle (L^*, K^*) is feasible (it produces at least y) but not necessarily cost-minimizing. Therefore:

$$c(w, r, y) \leq wL^* + rK^* \quad \text{for all } (w, r). \quad (18)$$

At $(w, r) = (\bar{w}, \bar{r})$, equality holds by definition:

$$c(\bar{w}, \bar{r}, y) = \bar{w}L^* + \bar{r}K^*. \quad (19)$$

Define $\phi(w, r) = wL^* + rK^* - c(w, r, y)$. By (18), $\phi(w, r) \geq 0$ for all (w, r) . By (19), $\phi(\bar{w}, \bar{r}) = 0$. Hence ϕ attains its minimum at $(\bar{w}, \bar{r})$. If c is differentiable at $(\bar{w}, \bar{r})$, the first-order condition $\nabla \phi(\bar{w}, \bar{r}) = 0$ gives:

$$\left(\frac{\partial c}{\partial w}, \frac{\partial c}{\partial r} \right) \Big|_{(\bar{w}, \bar{r})} = (L^*, K^*) = \mathbf{z}(\bar{w}, \bar{r}, y).$$

Part (v): The substitution matrix is symmetric NSD with $D^2c \cdot (\bar{w}, \bar{r})^\top = 0$.

Symmetry. By Shephard's lemma, $\partial c / \partial w = L(w, r, y)$ and $\partial c / \partial r = K(w, r, y)$. If c is twice continuously differentiable, Young's theorem (equality of cross-partial) gives:

$$\frac{\partial L}{\partial r} = \frac{\partial^2 c}{\partial r \partial w} = \frac{\partial^2 c}{\partial w \partial r} = \frac{\partial K}{\partial w}.$$

That is, $D_{(w,r)}^2 c$ is symmetric.

Negative semidefiniteness. Since c is concave in (w, r) (Part (ii)), its Hessian $D_{(w,r)}^2 c$ is negative semidefinite. To see this directly: for any vector $v = (\Delta w, \Delta r) \in \mathbb{R}^2$,

$$v^\top D_{(w,r)}^2 c v = \lim_{t \rightarrow 0} \frac{c(\bar{w} + t\Delta w, \bar{r} + t\Delta r, y) - 2c(\bar{w}, \bar{r}, y) + c(\bar{w} - t\Delta w, \bar{r} - t\Delta r, y)}{t^2} \leq 0,$$

where the inequality follows from concavity: $c(\bar{w} + t\Delta w, \bar{r} + t\Delta r, y) + c(\bar{w} - t\Delta w, \bar{r} - t\Delta r, y) \leq 2c(\bar{w}, \bar{r}, y)$ (applying the concavity definition with $\alpha = 1/2$).

In particular, the diagonal entries satisfy:

$$\frac{\partial L}{\partial w} = \frac{\partial^2 c}{\partial w^2} \leq 0, \quad \frac{\partial K}{\partial r} = \frac{\partial^2 c}{\partial r^2} \leq 0.$$

This confirms that conditional factor demands are downward-sloping in own price.

$D^2c \cdot (\bar{w}, \bar{r})^\top = 0$. By homogeneity of degree one (Part (i)), Euler's theorem applied to the first partial derivatives gives:

$$\bar{w} \frac{\partial^2 c}{\partial w^2} + \bar{r} \frac{\partial^2 c}{\partial r \partial w} = 0 \quad \text{and} \quad \bar{w} \frac{\partial^2 c}{\partial w \partial r} + \bar{r} \frac{\partial^2 c}{\partial r^2} = 0.$$

To derive this: since $c(tw, tr, y) = tc(w, r, y)$, differentiating with respect to w gives $t \partial c(tw, tr, y) / \partial (tw) = \partial c(w, r, y) / \partial w$, i.e., $\partial c / \partial w$ is Homog. of degree zero in (w, r) . By Euler's theorem for Homog.-of-degree-zero functions:

$$\bar{w} \frac{\partial}{\partial \bar{w}} \left(\frac{\partial c}{\partial w} \right) + \bar{r} \frac{\partial}{\partial \bar{r}} \left(\frac{\partial c}{\partial w} \right) = 0 \cdot \frac{\partial c}{\partial w} = 0.$$

Analogously for $\partial c / \partial r$. In matrix form, this is $D_{(w,r)}^2 c \cdot (\bar{w}, \bar{r})^\top = 0$.

Economically, this means that a proportional increase in all factor prices does not change the cost-minimizing input mix (since $\mathbf{z}$ is Homog. of degree zero in (w, r)), so the substitution effects along the direction $(\bar{w}, \bar{r})$ are zero.

Part (vi): CRS implies c and $\mathbf{z}$ are Homog. of degree one in y .

Suppose F is Homog. of degree one: $F(tL, tK) = tF(L, K)$ for all $t > 0$. Let (L^*, K^*) be cost-minimizing for output y :

$$c(w, r, y) = wL^* + rK^*, \quad F(L^*, K^*) = y.$$

For output ty , consider the bundle (tL^*, tK^*) . By CRS: $F(tL^*, tK^*) = tF(L^*, K^*) = ty$, so this bundle is feasible for output ty , and its cost is $t(wL^* + rK^*) = tc(w, r, y)$. Hence $c(w, r, ty) \leq tc(w, r, y)$.

For the reverse inequality, let (L', K') be cost-minimizing for output ty :

$$c(w, r, ty) = wL' + rK', \quad F(L', K') = ty.$$

Then $(L'/t, K'/t)$ produces $F(L'/t, K'/t) = (1/t)F(L', K') = y$ by CRS, at cost $(1/t)(wL' + rK') = (1/t)c(w, r, ty)$. Hence $c(w, r, y) \leq (1/t)c(w, r, ty)$, i.e., $tc(w, r, y) \leq c(w, r, ty)$.

Combining: $c(w, r, ty) = tc(w, r, y)$. Moreover, the argument shows $\mathbf{z}(w, r, ty) = t\mathbf{z}(w, r, y)$.

Economically, under CRS, doubling output exactly doubles cost. Average cost $c(w, r, y)/y$ is independent of y , and marginal cost $\partial c / \partial y$ equals average cost.

Part (vii): Concavity of F implies convexity of c in y .

Let $y^0, y^1 \geq 0$ and $\alpha \in [0, 1]$. Define $y^\alpha = \alpha y^0 + (1 - \alpha)y^1$. Let $\mathbf{x}^0 = (L^0, K^0) \in \mathbf{z}(w, r, y^0)$ and $\mathbf{x}^1 = (L^1, K^1) \in \mathbf{z}(w, r, y^1)$, and define $\mathbf{x}^\alpha = \alpha \mathbf{x}^0 + (1 - \alpha)\mathbf{x}^1$.

By concavity of F :

$$F(\mathbf{x}^\alpha) = F(\alpha \mathbf{x}^0 + (1 - \alpha)\mathbf{x}^1) \geq \alpha F(\mathbf{x}^0) + (1 - \alpha)F(\mathbf{x}^1) \geq \alpha y^0 + (1 - \alpha)y^1 = y^\alpha.$$

So $\mathbf{x}^\alpha$ is feasible for output y^α . Therefore:

$$\begin{aligned} c(w, r, y^\alpha) &\leq w(\alpha L^0 + (1 - \alpha)L^1) + r(\alpha K^0 + (1 - \alpha)K^1) \\ &= \alpha(wL^0 + rK^0) + (1 - \alpha)(wL^1 + rK^1) \\ &= \alpha c(w, r, y^0) + (1 - \alpha)c(w, r, y^1). \end{aligned}$$

Hence c is convex in y .

Since c is convex in y , the marginal cost $MC(y) = \partial c / \partial y$ is Increasing (a convex function has a Increasing derivative). Economically: as the firm expands output, each additional unit costs at least as much as the previous one, because the concavity of F means the firm faces diminishing returns.

Proof of (ii): Concavity in factor prices. Fix output at y . Let $\mathbf{z}$ minimize cost at w , and $\mathbf{z}'$ minimize cost at w' , and $\mathbf{z}''$ minimize cost at $w'' = tw + (1 - t)w'$. Then:

$$c(w'', y) = w'' \cdot \mathbf{z}'' = tw \cdot \mathbf{z}'' + (1 - t)w' \cdot \mathbf{z}''.$$

By cost minimization, $w \cdot \mathbf{z}'' \geq w \cdot \mathbf{z} = c(w, y)$ and $w' \cdot \mathbf{z}'' \geq w' \cdot \mathbf{z}' = c(w', y)$. Hence $c(w'', y) \geq tc(w, y) + (1 - t)c(w', y)$. $\square$

Economic intuition for concavity in w

Why is the cost function concave in factor prices? Consider the “passive” cost if the firm did not re-optimize when prices change: this would be linear in w . But the firm *does* re-optimize—when one factor becomes expensive, it substitutes toward the cheaper factor. So actual costs rise *less than linearly* when a factor price increases. This is exactly concavity. Compare with the profit function, which is *convex* in output prices for the same reason: the firm re-optimizes by producing more when prices rise, gaining *more than linearly*.

Remark 3.5. The parallel with consumer theory is striking: the cost function plays the role of the expenditure function, conditional factor demand corresponds to compensated (Hicksian) demand, and Shephard’s lemma is the analog of the derivative property of the expenditure function. We develop this parallel systematically in Section 6.

3.4.1 Revealed Preference in Cost Minimization

Just as profit maximization has observable implications (the law of supply), cost minimization has its own revealed-preference characterization that does *not* require differentiability or any parametric assumptions on the technology.

Definition 3.5 (Weak Axiom of Cost Minimization (WACM)). A set of observations $\{(w^t, \mathbf{x}^t, y^t)\}_{t=1}^T$ (factor prices, input bundles, and output levels) satisfies the *Weak Axiom of Cost Minimization* if for all t, s with $y^s \geq y^t$:

$$w^t \cdot \mathbf{x}^t \leq w^t \cdot \mathbf{x}^s.$$

That is, the observed input bundle at observation t must cost no more (at observation t ’s prices) than any other observed bundle that produces at least as much output.

Proposition 3.5 (Downward-sloping factor demands). *If observations t and s satisfy WACM and $y^t = y^s$, then*

$$\Delta w \cdot \Delta \mathbf{x} \leq 0, \tag{20}$$

where $\Delta w = w^t - w^s$ and $\Delta \mathbf{x} = \mathbf{x}^t - \mathbf{x}^s$.

Proof. WACM gives $w^t \cdot \mathbf{x}^t \leq w^t \cdot \mathbf{x}^s$ and $w^s \cdot \mathbf{x}^s \leq w^s \cdot \mathbf{x}^t$. Adding:

$$w^t \cdot (\mathbf{x}^t - \mathbf{x}^s) + w^s \cdot (\mathbf{x}^s - \mathbf{x}^t) \leq 0 \implies (w^t - w^s) \cdot (\mathbf{x}^t - \mathbf{x}^s) \leq 0. \quad \square$$

Remark 3.6. This is the cost-minimization analog of the law of supply $\Delta p \cdot \Delta y \geq 0$ for profit maximization. It says: when a factor becomes more expensive, the firm uses less of it (substituting toward cheaper factors). No assumptions on the technology (convexity, differentiability) are needed—just cost-minimizing behavior.

4 Other Topics

4.1 Income Distribution Under CRS

Theorem 4.1 (Euler's theorem and factor payments). *If F is Homog. of degree one (CRS) and differentiable, then*

$$F(L, K) = F_L(L, K) \cdot L + F_K(L, K) \cdot K. \quad (21)$$

Under profit maximization ($w = pF_L$ and $r = pF_K$), this gives

$$py = wL + rK. \quad (22)$$

Revenue is fully exhausted by factor payments—profit is zero.

Proof. Since $F(tL, tK) = tF(L, K)$, differentiate both sides with respect to t and evaluate at $t = 1$:

$$F_L(tL, tK) \cdot L + F_K(tL, tK) \cdot K \Big|_{t=1} = F(L, K). \quad \square$$

Source of income inequality

In a competitive market, the payment to each agent equals the value of the marginal product of the input owned by that agent:

$$\underbrace{\left(p \frac{\partial F}{\partial L}\right) L}_{\text{total payment to workers}} + \underbrace{\left(p \frac{\partial F}{\partial K}\right) K}_{\text{total payment to capital}} = \underbrace{py}_{\text{sale}}.$$

Income inequality arises from differences in marginal productivity. By diminishing returns, factors that are *scarce* command higher marginal products and hence higher returns.

4.2 International Productivity and Factor Price Comparisons

The first-order conditions for profit maximization link factor prices to marginal productivities. This connection has direct implications for international comparisons of wages and capital costs.

Proposition 4.1 (Factor prices reflect marginal productivities). *Suppose two countries, A and B , share the same CRS production function F with $F_{LK} > 0$, face the same output price $p > 0$, and both operate at interior profit-maximizing allocations. If country A has a higher labor-to-capital ratio than country B ($L^A/K^A > L^B/K^B$), then:*

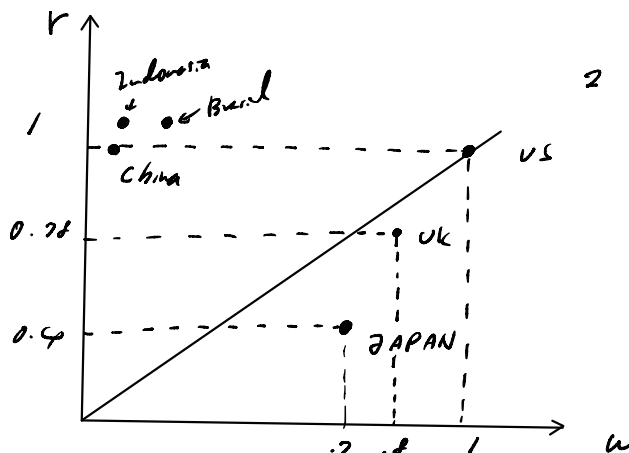
- (i) *Country A has a lower wage and higher rental price: $w^A < w^B$ and $r^A > r^B$.*
- (ii) *The wage-rental ratio is lower in the labor-abundant country: $w^A/r^A < w^B/r^B$.*

Proof. The profit-maximizing FOCs give $w = pF_L(L, K)$ and $r = pF_K(L, K)$, so

$$\frac{w}{r} = \frac{F_L(L, K)}{F_K(L, K)} = MRTS_{LK}(L, K). \quad (23)$$

Under CRS, F is Homog. of degree 1, so F_L and F_K are Homog. of degree 0. Therefore they depend only on the ratio $\ell = L/K$:

$$F_L(L, K) = F_L(\ell, 1) \equiv \phi(\ell), \quad F_K(L, K) = F_K(\ell, 1) \equiv \psi(\ell).$$



We now show that ϕ is decreasing and ψ is increasing in ℓ .

Differentiate Euler's identity $F(L, K) = F_L \cdot L + F_K \cdot K$ with respect to L :

$$F_L = F_{LL} \cdot L + F_L + F_{KL} \cdot K \implies F_{LL} \cdot L + F_{KL} \cdot K = 0.$$

Since $F_{KL} > 0$ and $K > 0$, we obtain $F_{LL} < 0$. Therefore $\phi'(\ell) = F_{LL}(\ell, 1) < 0$: the marginal product of labor is decreasing in the labor-capital ratio. An analogous argument (differentiating Euler's identity with respect to K) gives $F_{KK} < 0$ and $\psi'(\ell) > 0$.

Since $\ell^A = L^A/K^A > L^B/K^B = \ell^B$:

$$w^A = p\phi(\ell^A) < p\phi(\ell^B) = w^B \quad \text{and} \quad r^A = p\psi(\ell^A) > p\psi(\ell^B) = r^B.$$

For part (ii), $w/r = \phi(\ell)/\psi(\ell)$ is the ratio of a decreasing function to an increasing function, hence strictly decreasing in ℓ . $\square$

Scarcity determines factor returns

A factor that is *scarce* relative to other factors commands a higher marginal product and therefore a higher price. In a labor-abundant country (high L/K), labor is cheap and capital is expensive. In a capital-abundant country (low L/K), the reverse holds. This is the theoretical foundation of the *Heckscher-Ohlin* prediction in international trade: countries export goods that use their abundant factor intensively.

Example 4.1 (Cobb–Douglas illustration). For $F(L, K) = L^\alpha K^{1-\alpha}$ with $0 < \alpha < 1$:

$$w = p\alpha \left(\frac{K}{L}\right)^{1-\alpha}, \quad r = p(1-\alpha) \left(\frac{L}{K}\right)^\alpha, \quad \frac{w}{r} = \frac{\alpha}{1-\alpha} \cdot \frac{K}{L}.$$

As L/K increases, w falls as $(K/L)^{1-\alpha}$ and r rises as $(L/K)^\alpha$.

4.3 Recovering the Production Function from the Profit Function

A key result of duality theory is that the production function can be recovered from the profit function. The mathematical tool is the Legendre–Fenchel transform (conjugate function).

4.3.1 The Legendre–Fenchel Transform

Definition 4.1 (Convex conjugate). Let $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function. The *convex conjugate* (or *Legendre–Fenchel transform*) of g is the function $g^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$g^*(p) = \sup_{x \in \mathbb{R}^n} [p \cdot x - g(x)].$$

Proposition 4.2 (Properties of the conjugate).

- (i) g^* is always convex (as a supremum of affine functions), regardless of whether g is convex.
- (ii) **Fenchel's inequality:** For all x and p ,

$$g(x) + g^*(p) \geq p \cdot x. \quad (24)$$

- (iii) Equality holds in (24) if and only if $p \in \partial g(x)$ (the subdifferential of g at x). If g is differentiable, this becomes $p = \nabla g(x)$.

Proof. Part (i): For each fixed x , the function $p \mapsto p \cdot x - g(x)$ is affine in p . The supremum over x of affine functions is convex.

Part (ii): By definition, $g^*(p) = \sup_z [p \cdot z - g(z)] \geq p \cdot x - g(x)$ for any particular x . Rearranging gives $g(x) + g^*(p) \geq p \cdot x$. $\square$

Theorem 4.2 (Biconjugation). If $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous, then

$$g^{**}(x) \equiv \sup_p [p \cdot x - g^*(p)] = g(x).$$

That is, the conjugate of the conjugate recovers the original function.

Geometric interpretation

The conjugate $g^*(p)$ encodes g through its supporting hyperplanes. For each slope p , $g^*(p)$ is the maximal gap between the linear function $x \mapsto p \cdot x$ and $g(x)$ —equivalently, $-g^*(p)$ is the y -intercept of the supporting hyperplane with slope p . The biconjugation theorem says that a convex function is completely determined by the family of its supporting hyperplanes.

4.3.2 Application to Production and Profit

Proposition 4.3 (Profit function as a conjugate). Let f be concave and let $g = -f$ (convex). Define the real wage $\omega = w/p$. The maximized real profit (profit per unit of output price) is

$$\tilde{\pi}(\omega) = \max_{L \geq 0} [f(L) - \omega L].$$

Then:

- (i) $\tilde{\pi}(\omega) = -g^*(-\omega)$, where g^* is the conjugate of $g = -f$.
- (ii) $\tilde{\pi}$ is convex and Decreasing in ω .
- (iii) By the biconjugation theorem, the production function can be recovered:

$$f(L) = \min_{\omega \geq 0} [\omega L + \tilde{\pi}(\omega)]. \quad (25)$$

Proof. Part (i): By definition of the conjugate,

$$g^*(p) = \sup_L [pL - g(L)] = \sup_L [pL + f(L)].$$

Setting $p = -\omega$:

$$g^*(-\omega) = \sup_L [-\omega L + f(L)] = \tilde{\pi}(\omega),$$

so $\tilde{\pi}(\omega) = g^*(-\omega)$.

Part (ii): g^* is convex by Proposition 4.2(i). The composition $\omega \mapsto g^*(-\omega)$ preserves convexity (since $\omega \mapsto -\omega$ is affine). Monotonicity: if ω increases, the objective $f(L) - \omega L$ decreases pointwise, so the maximum decreases.

Part (iii): By the biconjugation theorem applied to $g = -f$:

$$g(L) = \sup_p [pL - g^*(p)] = \sup_p [pL - g^*(p)].$$

Substituting $g = -f$ and $p = -\omega$:

$$-f(L) = \sup_{\omega} [-\omega L - \tilde{\pi}(\omega)] = -\inf_{\omega} [\omega L + \tilde{\pi}(\omega)].$$

Hence $f(L) = \inf_{\omega \geq 0} [\omega L + \tilde{\pi}(\omega)]$. (The restriction $\omega \geq 0$ is without loss since $\tilde{\pi}(\omega) = +\infty$ for $\omega < 0$ when f is unbounded.) $\square$

Example 4.2 (Recovery for Cobb–Douglas). Let $f(L) = L^{1/2}$ and $w = p = 1$ (so $\omega = 1$). The real profit is $\tilde{\pi}(\omega) = \max_L [L^{1/2} - \omega L]$. The FOC $\frac{1}{2}L^{-1/2} = \omega$ gives $L^* = 1/(4\omega^2)$, so

$$\tilde{\pi}(\omega) = \frac{1}{2\omega} - \frac{1}{4\omega} = \frac{1}{4\omega}.$$

Applying (25):

$$f(L) = \min_{\omega > 0} \left[\omega L + \frac{1}{4\omega} \right].$$

The FOC $L - 1/(4\omega^2) = 0$ gives $\omega^* = 1/(2\sqrt{L})$, and substituting back:

$$f(L) = \frac{\sqrt{L}}{2} + \frac{\sqrt{L}}{2} = \sqrt{L} = L^{1/2},$$

recovering the original production function.

Remark 4.1 (Parallel with consumer theory). This duality mirrors the consumer-theory identity: the indirect utility function $v(p, w)$ and the expenditure function $e(p, u)$ are related by conjugation, and the utility function can be recovered from either via $u(x) = \min_p [e(p, u) : p \cdot x \leq w]$. In production, the profit function $\pi(p, w)$ plays the role of the indirect utility, and the production function f plays the role of the direct utility. See Section 6 for the full parallel.

4.4 Applicability of Production Theory to Discrete Data

The theory developed above assumes:

- (a) **Divisibility:** inputs can be varied continuously.

- (b) **Convexity:** the input requirement set $V(y) = \{(L, K) : F(L, K) \geq y\}$ is convex, i.e., isoquants are convex to the origin.

In practice, inputs are often indivisible (you cannot hire half a truck or a third of a worker), and production possibilities may be non-convex. The following proposition shows that convexity provides a useful approximation even in these cases.

Proposition 4.4 (Convexification of discrete technologies). *Let $\mathcal{T} = \{(L_1, K_1), \dots, (L_m, K_m)\} \subset \mathbb{R}_+^2$ be a finite set of input bundles, each capable of producing at least $\bar{y}$ units of output. Define:*

- (i) *The discrete input requirement set: $V_{disc}(\bar{y}) = \mathcal{T} + \mathbb{R}_+^2$ (the set $\mathcal{T}$ plus free disposal).*
- (ii) *The convexified input requirement set: $V_{conv}(\bar{y}) = \text{conv}(\mathcal{T}) + \mathbb{R}_+^2$ (the convex hull of $\mathcal{T}$ plus free disposal).*

Then:

- (i) $V_{disc}(\bar{y}) \subseteq V_{conv}(\bar{y})$.
- (ii) *The cost function is the same for both:*

$$\min_{\mathbf{x} \in V_{disc}(\bar{y})} w \cdot \mathbf{x} = \min_{\mathbf{x} \in V_{conv}(\bar{y})} w \cdot \mathbf{x}$$

for all $w \geq 0$ such that the minimum on the left is attained at a vertex of $\text{conv}(\mathcal{T})$.

- (iii) *More generally, if there are J firms each with a (possibly non-convex) technology $V_j(\bar{y})$, the aggregate input requirement set $\frac{1}{J} \sum_{j=1}^J V_j(\bar{y})$ approaches a convex set as $J \rightarrow \infty$ (Shapley–Folkman theorem).*

Proof of (ii). Let $\mathbf{x}^*$ solve $\min_{\mathbf{x} \in V_{conv}(\bar{y})} w \cdot \mathbf{x}$. Since $V_{conv}(\bar{y})$ is a convex polytope (plus $\mathbb{R}_+^2$), the minimum is attained at a vertex by the fundamental theorem of linear programming. The vertices of $\text{conv}(\mathcal{T}) + \mathbb{R}_+^2$ that can be cost-minimizing are exactly the points in $\mathcal{T}$ (the extreme points of the convex hull, possibly shifted by free disposal). Hence $\mathbf{x}^*$ can be chosen in $V_{disc}(\bar{y})$, and the two minima coincide. $\square$

Example 4.3 (Moving company). A moving company can transport 10 tons using the following discrete input combinations (L, K) (workers, trucks):

$$\mathcal{T} = \{(6, 1), (4, 2), (3, 3), (2, 5), (1, 8)\}.$$

The discrete input requirement set $V_{disc}(10)$ consists of these five points plus anything with weakly more of both inputs (free disposal). The isoquant is a set of five isolated points.

The convexified isoquant $\text{conv}(\mathcal{T})$ is the piecewise-linear curve connecting these points, forming the lower-left boundary of $V_{conv}(10)$. Along this convex boundary, the MRTS is well-defined and decreasing (on each linear segment), and cost minimization gives a unique tangency with the iso-cost line.

Although the firm cannot actually operate at, say, $(3.5, 2.5)$, the convexified technology gives the correct cost function at any factor prices where the cost-minimizing bundle happens to be one of the five actual input combinations. With many firms operating at different points, the aggregate technology is approximately convex even if each individual firm faces discrete choices.

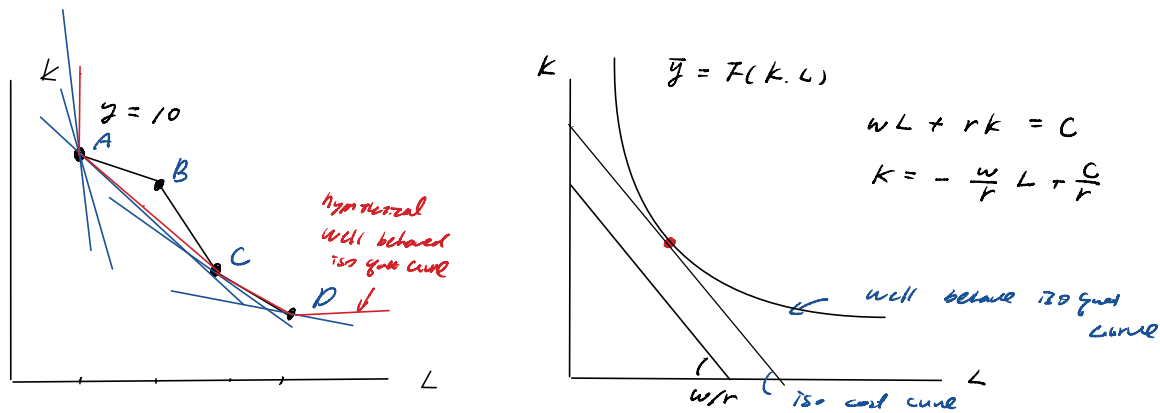


Figure 5: Discrete input combinations (dots) and their convex hull (dashed curve). The convexified isoquant provides a smooth approximation of the firm's technology. By the Shapley–Folkman theorem, the aggregate technology of many such firms is approximately convex.

Why convexity is a reasonable approximation

There are three justifications for using convex technologies even when individual production is discrete or non-convex:

- (i) **Linear programming:** Cost minimization over a discrete technology is equivalent to cost minimization over its convex hull (the minima coincide at vertices).
- (ii) **Aggregation:** With many firms, the average technology becomes approximately convex (Shapley–Folkman). Non-convexities “wash out” in aggregation.
- (iii) **Time averaging:** Even a single firm can approximate a convex combination by time-sharing between techniques (e.g., using technique A for half the month and technique B for the other half).

5 General Production Sets

We now present the general framework following Mas-Colell, Whinston, and Green (1995), Chapter 5.

5.1 Production Sets and Transformation Functions

Definition 5.1 (Production set). Consider an economy with L commodities. A *production vector* (or *netput vector*) is $y = (y_1, \dots, y_L) \in \mathbb{R}^L$, where $y_\ell > 0$ denotes a net output and $y_\ell < 0$ denotes a net input. The *production set* $Y \subset \mathbb{R}^L$ is the set of all technologically feasible production vectors. Any $y \in Y$ is feasible; $y \notin Y$ is not.

Definition 5.2 (Transformation function). The *transformation function* $T : \mathbb{R}^L \rightarrow \mathbb{R}$ describes Y via

$$Y = \{y \in \mathbb{R}^L : T(y) \leq 0\},$$

with $T(y) = 0$ on the boundary of Y , called the *transformation frontier*. If T is differentiable and $T(\bar{y}) = 0$, the *marginal rate of transformation* of good ℓ for good k at $\bar{y}$ is

$$MRT_{\ell k}(\bar{y}) = \frac{\partial T(\bar{y})/\partial y_\ell}{\partial T(\bar{y})/\partial y_k}.$$

Remark 5.1 (Relationship to the production function). The transformation function T generalizes the production function f from Sections 2–3 in two ways.

(i) **Single output.** Consider a firm that uses n inputs $(x_1, \dots, x_n)$ to produce a single output q , with production function $q = f(x_1, \dots, x_n)$. In the netput convention, inputs enter negatively: the production vector is $y = (q, -x_1, \dots, -x_n) \in \mathbb{R}^{n+1}$, where $y_1 = q > 0$ is the output and $y_i = -x_{i-1} < 0$ for $i = 2, \dots, n+1$ are the inputs. The transformation function is:

$$T(y) = y_1 - F(-y_2, \dots, -y_{n+1}).$$

Then $T(y) \leq 0$ if and only if $q \leq F(x_1, \dots, x_n)$, which is exactly the condition that $(q, x_1, \dots, x_n)$ is feasible. The transformation frontier $T(y) = 0$ corresponds to efficient production $q = F(x_1, \dots, x_n)$.

For example, with a single input (L) and single output (q), the netput vector is $y = (q, -L)$ and $T(q, -L) = q - f(L)$. The production set is:

$$Y = \{(q, -L) : q \leq f(L), L \geq 0\} = \{(q, -L) : T(q, -L) \leq 0\}.$$

(ii) **Multiple outputs.** The transformation function handles multiple outputs naturally, which the production function cannot. Consider a firm producing two outputs (y_1, y_2) from a single input x , with netput $y = (y_1, y_2, -x)$. The transformation function $T(y_1, y_2, -x) \leq 0$ describes the production possibilities frontier (PPF) for the two outputs given input x . There is no natural way to write this as a “production function” mapping inputs to outputs, since the output is multi-dimensional.

(iii) **The MRT generalizes the MRTS.** In the single-output case with inputs x_1, x_2 , the netput vector is $y = (q, -x_1, -x_2)$ and $T(y) = q - f(x_1, x_2)$. Then:

$$MRT_{2,3}(y) = \frac{\partial T/\partial y_2}{\partial T/\partial y_3} = \frac{-\partial f/\partial x_1}{-\partial f/\partial x_2} = \frac{f_1}{f_2} = MRTS_{12}.$$

For two outputs y_1, y_2 (holding inputs fixed), $MRT_{1,2}$ measures the rate at which the firm must sacrifice output 2 to produce more of output 1—the slope of the PPF.

Example 5.1 (From production function to transformation function).

- (i) **Cobb–Douglas, single output:** $f(L, K) = L^\alpha K^\beta$. Netput $y = (q, -L, -K)$.

$$T(y_1, y_2, y_3) = y_1 - (-y_2)^\alpha (-y_3)^\beta.$$

The production set is $Y = \{(q, -L, -K) : q \leq L^\alpha K^\beta\}$.

- (ii) **Linear, two outputs:** A firm uses labor L to produce goods 1 and 2, with a units of labor per unit of good 1 and b units per unit of good 2. Netput $y = (y_1, y_2, -L)$.

$$T(y_1, y_2, y_3) = ay_1 + by_2 + y_3,$$

so $Y = \{(y_1, y_2, -L) : ay_1 + by_2 \leq L\}$. The transformation frontier $ay_1 + by_2 = L$ is the familiar linear PPF with slope $-a/b$.

Summary of the conceptual hierarchy

The three representations of technology are related as follows:

$$\underbrace{F(x_1, \dots, x_n)}_{\text{Production function}} \longrightarrow \underbrace{Y \subset \mathbb{R}^L}_{\text{Production set}} \longleftrightarrow \underbrace{T(y) \leq 0}_{\text{Transformation function}}$$

The production function F is the most restrictive: it assumes a single output and describes the efficient boundary $q = f(\mathbf{x})$. The production set Y is the most general: it is simply a subset of $\mathbb{R}^L$ describing all feasible netput vectors. The transformation function T is a convenient analytical description of Y that allows calculus (FOCs, envelope theorems) to be applied in the fully general multi-input, multi-output setting. In the single-output case, all three are equivalent; with multiple outputs, only Y and T apply.

5.2 Properties of Production Sets

The following is a catalog of commonly assumed properties:

- (i) **Y is nonempty.** The firm has something it can do.
- (ii) **Y is closed.** Limits of feasible production plans are feasible.
- (iii) **No free lunch.** $Y \cap \mathbb{R}_+^L \subseteq \{0\}$. One cannot produce output without inputs.
- (iv) **Possibility of inaction.** $0 \in Y$. The firm can choose to do nothing.
- (v) **Free disposal.** $Y - \mathbb{R}_+^L \subseteq Y$. Extra inputs (or outputs) can be disposed of at no cost.
- (vi) **Irreversibility.** If $y \in Y$ and $y \neq 0$, then $-y \notin Y$.
- (vii) **Decreasing returns to scale.** If $y \in Y$, then $\alpha y \in Y$ for all $\alpha \in [0, 1]$.
- (viii) **Increasing returns to scale.** If $y \in Y$, then $\alpha y \in Y$ for all $\alpha \geq 1$.
- (ix) **Constant returns to scale.** If $y \in Y$, then $\alpha y \in Y$ for all $\alpha \geq 0$. (Geometrically, Y is a cone.)

(x) **Additivity (free entry).** If $y, y' \in Y$, then $y + y' \in Y$, i.e., $Y + Y \subseteq Y$.

(xi) **Convexity.** If $y, y' \in Y$ and $\alpha \in [0, 1]$, then $\alpha y + (1 - \alpha)y' \in Y$.

(xii) **Y is a convex cone.** Conjunction of (ix) and (xi).

Proposition 5.1 (Additivity and convexity). *Y is additive and satisfies Decreasing returns if and only if Y is a convex cone.*

Proof. ($\Leftarrow$): A convex cone is additive (since $y + y' = 2[\frac{1}{2}y + \frac{1}{2}y']$ and scaling) and satisfies Decreasing returns by definition.

($\Rightarrow$): For any $y, y' \in Y$ and $\alpha > 0$, $\beta > 0$, we need $\alpha y + \beta y' \in Y$. Let $k > \max\{\alpha, \beta\}$ be an integer. By additivity, $ky \in Y$. Since $\alpha/k < 1$, Decreasing returns gives $\alpha y = (\alpha/k)(ky) \in Y$. Similarly $\beta y' \in Y$. By additivity, $\alpha y + \beta y' \in Y$. $\square$

Proposition 5.2 (CRS and convex production). *For any convex production set $Y \subset \mathbb{R}^L$ with $0 \in Y$, there exists a constant returns, convex set $Y' \subset \mathbb{R}^{L+1}$ such that $Y' = \{y \in \mathbb{R}^{L+1} : (y, -1) \in Y' \text{ for some scaling}\}$.*

Remark 5.2. This shows that, in a competitive convex setting, there is little loss of generality in assuming CRS—one can always add an “entrepreneurial input” to convert DRS into CRS in a higher-dimensional space.

5.3 Profit Maximization

Given a price vector $p \in \mathbb{R}_{++}^L$, the *profit maximization problem (PMP)* is:

$$\pi(p) = \max_{y \in Y} p \cdot y. \quad (26)$$

The *profit function* $\pi(p)$ gives the value of the solution. The *supply correspondence* $y(p) = \{y \in Y : p \cdot y = \pi(p)\}$ gives the set of optimizers.

Proposition 5.3 (Properties of $\pi(\cdot)$ and $y(\cdot)$). *Suppose Y is closed, and satisfies free disposal. Then:*

- (i) $\pi(\cdot)$ is Homog. of degree one in p .
- (ii) $\pi(\cdot)$ is convex.
- (iii) If Y is convex, then $Y = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \text{ for all } p \gg 0\}$.
- (iv) $y(\cdot)$ is Homog. of degree zero in p .
- (v) If Y is convex, then $y(p)$ is a convex set for all p . If Y is strictly convex, $y(p)$ is single-valued.
- (vi) (Hotelling’s lemma) If $y(\bar{p})$ is a singleton and π is differentiable at $\bar{p}$, then $\nabla \pi(\bar{p}) = y(\bar{p})$.
- (vii) $D_p y(\bar{p}) = D^2 \pi(\bar{p})$ is a symmetric, positive semidefinite matrix.

Proof of convexity of π . For $\alpha \in [0, 1]$:

$$\begin{aligned}\pi(\alpha p + (1 - \alpha)p') &= \max_{y \in Y} [\alpha p + (1 - \alpha)p'] \cdot y \\ &= \max_{y \in Y} [\alpha(p \cdot y) + (1 - \alpha)(p' \cdot y)] \\ &\leq \alpha \max_{y \in Y} (p \cdot y) + (1 - \alpha) \max_{y \in Y} (p' \cdot y) \\ &= \alpha \pi(p) + (1 - \alpha) \pi(p').\end{aligned}$$

□

Proof. Throughout, the profit maximization problem is

$$\pi(p) = \sup_{y \in Y} p \cdot y, \quad y(p) = \{y \in Y : p \cdot y = \pi(p)\}.$$

Part (i): π is Homog. of degree one in p .

For any $t > 0$:

$$\pi(tp) = \sup_{y \in Y} (tp) \cdot y = t \sup_{y \in Y} p \cdot y = t \pi(p).$$

The production set Y does not depend on p , so scaling prices simply scales the objective.

Part (ii): π is convex.

Let $\alpha \in [0, 1]$ and $p^\alpha = \alpha p^0 + (1 - \alpha)p^1$. For any fixed $y \in Y$:

$$p^\alpha \cdot y = \alpha(p^0 \cdot y) + (1 - \alpha)(p^1 \cdot y) \leq \alpha \sup_{y' \in Y} p^0 \cdot y' + (1 - \alpha) \sup_{y' \in Y} p^1 \cdot y' = \alpha \pi(p^0) + (1 - \alpha) \pi(p^1).$$

Since this holds for every $y \in Y$, taking the supremum over y on the left:

$$\pi(p^\alpha) = \sup_{y \in Y} p^\alpha \cdot y \leq \alpha \pi(p^0) + (1 - \alpha) \pi(p^1).$$

□

Part (iii): If Y is convex, then $Y = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \text{ for all } p \gg 0\}$.

Define $Y^* = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \text{ for all } p \gg 0\}$.

$Y \subseteq Y^*$: If $y \in Y$, then for every $p \gg 0$, $p \cdot y \leq \sup_{y' \in Y} p \cdot y' = \pi(p)$ by definition of π . Hence $y \in Y^*$.

$Y^* \subseteq Y$: Suppose $\hat{y} \notin Y$. Since Y is closed and convex, the strict separating hyperplane theorem guarantees the existence of $\bar{p} \in \mathbb{R}^L$ and $\varepsilon > 0$ such that

$$\bar{p} \cdot \hat{y} > \bar{p} \cdot y + \varepsilon \quad \text{for all } y \in Y.$$

In particular, $\bar{p} \cdot \hat{y} > \sup_{y \in Y} \bar{p} \cdot y = \pi(\bar{p})$.

We need to show we can choose a strictly positive price vector. By free disposal, if $y \in Y$ then $y - te_\ell \in Y$ for all $t \geq 0$ and any unit vector e_ℓ . If $\bar{p}_\ell < 0$ for some ℓ , then $\bar{p} \cdot (y - te_\ell) = \bar{p} \cdot y - t\bar{p}_\ell \rightarrow +\infty$ as $t \rightarrow +\infty$, which would make $\pi(\bar{p}) = +\infty$, contradicting $\bar{p} \cdot \hat{y} > \pi(\bar{p})$. Hence $\bar{p} \geq 0$. A perturbation argument (replacing $\bar{p}$ by $\bar{p} + \delta \mathbf{1}$ for small $\delta > 0$) yields a $p \gg 0$ with $p \cdot \hat{y} > \pi(p)$. Hence $\hat{y} \notin Y^*$.

Part (iv): $y(\cdot)$ is Homog. of degree zero in p .

For $t > 0$:

$$y(tp) = \{y \in Y : (tp) \cdot y = \pi(tp)\} = \{y \in Y : t(p \cdot y) = t \pi(p)\} = \{y \in Y : p \cdot y = \pi(p)\} = y(p).$$

The second equality uses Part (i). Scaling all prices does not change which production plan maximizes profit.

Part (v): Convexity of Y implies $y(p)$ is convex; strict convexity implies $y(p)$ is a singleton.

Convexity of $y(p)$: Let $y^0, y^1 \in y(p)$, so $p \cdot y^0 = p \cdot y^1 = \pi(p)$ and $y^0, y^1 \in Y$. For $\alpha \in [0, 1]$, let $y^\alpha = \alpha y^0 + (1 - \alpha)y^1$. Since Y is convex, $y^\alpha \in Y$. Moreover:

$$p \cdot y^\alpha = \alpha(p \cdot y^0) + (1 - \alpha)(p \cdot y^1) = \alpha \pi(p) + (1 - \alpha) \pi(p) = \pi(p).$$

Hence $y^\alpha \in y(p)$.

Strict convexity implies singleton: Suppose $y^0, y^1 \in y(p)$ with $y^0 \neq y^1$. Since Y is strictly convex, $y^{1/2} = \frac{1}{2}y^0 + \frac{1}{2}y^1$ lies in the interior of Y . By free disposal, there exists $\tilde{y} \in Y$ with $\tilde{y} \gg y^{1/2}$. Since $p \gg 0$:

$$p \cdot \tilde{y} > p \cdot y^{1/2} = \frac{1}{2}(p \cdot y^0) + \frac{1}{2}(p \cdot y^1) = \pi(p),$$

contradicting the definition of $\pi(p)$ as the supremum. Hence $y^0 = y^1$.

Part (vi): Hotelling's lemma.

We give two proofs.

Proof via the envelope theorem. The profit maximization problem $\pi(p) = \max_{y \in Y} p \cdot y$ has objective $p \cdot y$ that is linear in the parameter p . By the envelope theorem:

$$\left. \frac{\partial \pi}{\partial p_\ell} \right|_{\bar{p}} = \left. \frac{\partial (p \cdot y)}{\partial p_\ell} \right|_{y=y^*} = y_\ell^*,$$

where $y^* = y(\bar{p})$. Hence $\nabla_p \pi(\bar{p}) = y(\bar{p})$.

Proof from first principles. Let $y^* = y(\bar{p})$ (a singleton by assumption). For any price vector p , the plan y^* is feasible but not necessarily profit-maximizing at p , so:

$$\pi(p) \geq p \cdot y^* \quad \text{for all } p. \quad (27)$$

At $p = \bar{p}$, equality holds: $\pi(\bar{p}) = \bar{p} \cdot y^*$.

Define $\phi(p) = \pi(p) - p \cdot y^*$. By (27), $\phi(p) \geq 0$ for all p , and $\phi(\bar{p}) = 0$. Hence ϕ attains its global minimum at $\bar{p}$. If π is differentiable at $\bar{p}$, the first-order necessary condition $\nabla \phi(\bar{p}) = 0$ gives:

$$\nabla_p \pi(\bar{p}) = y^*.$$

Remark 5.3. The structure of the proof is identical to that of Shephard's lemma (Part (iv) of Proposition 3.4), but with the inequality reversed: in cost minimization, the cost function lies *below* the linear function $w \cdot \mathbf{x}^*$ (giving concavity), while in profit maximization, the profit function lies *above* the linear function $p \cdot y^*$ (giving convexity). This sign difference reflects the fact that cost is minimized (concave envelope) while profit is maximized (convex envelope).

Part (vii): $D_p y(\bar{p}) = D^2 \pi(\bar{p})$ is symmetric and positive semidefinite.

Symmetry. By Hotelling's lemma, $y_\ell(p) = \partial \pi / \partial p_\ell$. If π is twice continuously differentiable, Young's theorem gives:

$$\frac{\partial y_\ell}{\partial p_k} = \frac{\partial^2 \pi}{\partial p_k \partial p_\ell} = \frac{\partial^2 \pi}{\partial p_\ell \partial p_k} = \frac{\partial y_k}{\partial p_\ell}.$$

Hence $D_p y(\bar{p}) = D^2 \pi(\bar{p})$ is symmetric.

Positive semidefiniteness. Since π is convex (Part (ii)), its Hessian is positive semidefinite. Explicitly: for any vector $v = (\Delta p_1, \dots, \Delta p_L) \in \mathbb{R}^L$ and $t > 0$, convexity of π gives:

$$\frac{1}{2}\pi(\bar{p} + tv) + \frac{1}{2}\pi(\bar{p} - tv) \geq \pi(\bar{p}).$$

Rearranging and taking the limit:

$$v^\top D^2\pi(\bar{p})v = \lim_{t \rightarrow 0} \frac{\pi(\bar{p} + tv) - 2\pi(\bar{p}) + \pi(\bar{p} - tv)}{t^2} \geq 0.$$

The diagonal entries have direct economic content:

$$\frac{\partial y_\ell}{\partial p_\ell} = \frac{\partial^2 \pi}{\partial p_\ell^2} \geq 0.$$

- If good ℓ is an *output* ($y_\ell > 0$): its supply is Increasing in own price (*law of supply*).
- If good ℓ is an *input* ($y_\ell < 0$): writing $y_\ell = -x_\ell$, we get $\partial(-x_\ell)/\partial p_\ell \geq 0$, i.e., $\partial x_\ell/\partial p_\ell \leq 0$. Input demand is Decreasing in own price (*law of demand for inputs*).

These are the component-by-component consequences of the law of supply $\Delta p \cdot \Delta y \geq 0$ (Theorem 5.1).

Contrast with consumer theory. The Hessian $D^2\pi$ is positive semidefinite *unconditionally*—no Slutsky compensation is needed. In consumer theory, the analogous matrix (the Slutsky substitution matrix) is negative semidefinite, but it requires subtracting the income effect. The difference arises because the firm has no budget constraint: there is no wealth effect to correct for.

5.4 The Law of Supply

Theorem 5.1 (Law of supply). *Let $y^0 \in y(p^0)$ and $y^1 \in y(p^1)$. Then*

$$(p^1 - p^0) \cdot (y^1 - y^0) \geq 0. \quad (28)$$

Proof. By profit maximization: $p^1 \cdot y^1 \geq p^1 \cdot y^0$ and $p^0 \cdot y^0 \geq p^0 \cdot y^1$. Adding:

$$p^1 \cdot y^1 - p^0 \cdot y^1 \geq p^1 \cdot y^0 - p^0 \cdot y^0 \implies (p^1 - p^0) \cdot (y^1 - y^0) \geq 0.$$

□

Corollary 5.1. *If only the price of good k changes ($p_\ell^1 = p_\ell^0$ for $\ell \neq k$), then*

$$(p_k^1 - p_k^0)(y_k^1 - y_k^0) \geq 0.$$

- If $y_k > 0$ (output): supply is Increasing in own price. (Law of supply)
- If $y_k < 0$ (input): demand for the input is Decreasing in its price. (Law of demand for inputs)

Notice that the above rules are derived only from profit maximization.

No compensation needed

Unlike consumer theory, the law of supply requires no compensation (no “Slutsky” correction). There is no wealth effect because there is no budget constraint—only the optimization $\max_{y \in Y} p \cdot y$.

5.5 The Weak Axiom of Profit Maximization (WAPM)

The law of supply was derived from the first-order conditions of profit maximization. But we can also derive testable implications *directly from observed data*, without assuming anything about the technology.

Definition 5.3 (WAPM). A set of observations $\{(p^t, y^t)\}_{t=1}^T$ of prices and production plans satisfies the *Weak Axiom of Profit Maximization* if for all $t, s = 1, \dots, T$:

$$p^t \cdot y^t \geq p^t \cdot y^s. \quad (29)$$

That is, the production plan chosen at prices p^t must yield at least as much profit at p^t as any other observed production plan.

Proposition 5.4 (WAPM implies the law of supply). *If WAPM holds for observations t and s , then $\Delta p \cdot \Delta y \geq 0$ where $\Delta p = p^t - p^s$ and $\Delta y = y^t - y^s$.*

Proof. WAPM gives $p^t \cdot y^t \geq p^t \cdot y^s$ and $p^s \cdot y^s \geq p^s \cdot y^t$. Adding: $(p^t - p^s) \cdot (y^t - y^s) \geq 0$. $\square$

Remark 5.4. The power of WAPM is that it applies to *discrete, finite* data with no assumptions on the technology (no differentiability, no convexity, no functional form). It is a pure revealed-preference condition.

5.6 Recoverability of the Production Set

A natural question: does WAPM exhaust all the implications of profit maximization, or are there other testable restrictions? The answer is that WAPM is *both necessary and sufficient*. This can be shown by constructing a production set that rationalizes any data satisfying WAPM.

Definition 5.4 (Inner and outer bounds). Given data $\{(p^t, y^t)\}_{t=1}^T$ satisfying WAPM:

- (i) The *inner bound* Y_I is the smallest convex, monotonic set containing all observed production plans:

$$Y_I = \text{convex, monotonic hull of } \{y^1, \dots, y^T\}.$$

- (ii) The *outer bound* Y_O is

$$Y_O = \{y \in \mathbb{R}^L : p^t \cdot y \leq p^t \cdot y^t \text{ for all } t = 1, \dots, T\}.$$

Proposition 5.5 (Rationalizability). *If $\{(p^t, y^t)\}_{t=1}^T$ satisfies WAPM, then:*

- (i) $Y_I \subseteq Y \subseteq Y_O$ for any convex, monotonic production set Y that could have generated the data.
- (ii) Both Y_I and Y_O rationalize the data: for each t , y^t is profit-maximizing at prices p^t in both Y_I and Y_O .

Proof sketch. For Y_I : If y^t did not maximize $p^t \cdot y$ over Y_I , there would exist some y^s (or convex combination) in Y_I with $p^t \cdot y^s > p^t \cdot y^t$, contradicting WAPM.

For Y_O : If $y \in Y_O$, then $p^t \cdot y \leq p^t \cdot y^t$ for all t , so y^t maximizes $p^t \cdot y$ over Y_O by construction. $\square$

What the bounds tell us

Y_I is the most conservative estimate of the technology: it includes only what we can be sure is feasible (observed points and their convex, monotonic combinations). Y_O is the most generous: it includes everything that is *not ruled out* by the observations. The true technology Y must lie between them. As we observe more data, the gap between Y_I and Y_O narrows.

6 Comparison: Producer Theory and Consumer Theory

Production theory and consumer theory have a deep structural parallel. Both involve optimizing agents, but the differences in the structure of the optimization problem lead to important asymmetries. The following table summarizes the correspondence.

Why producer theory is simpler

The fundamental reason producer theory is simpler is the *absence of a budget constraint*. In consumer theory, the budget constraint creates a wealth effect: when prices change, the consumer's real purchasing power changes, leading to the Slutsky decomposition into substitution and income effects. In producer theory, there is no wealth to redistribute—the firm simply maximizes $p \cdot y$ over Y . This means:

- (i) The law of supply holds *unconditionally* (no compensation needed).
- (ii) The profit function is convex (not merely quasiconvex).
- (iii) Net supply functions satisfy $D_p y = D^2 \pi \succeq 0$ (PSD), not just “NSD after Slutsky correction.”
- (iv) The substitution matrix for net supply is always symmetric PSD, with no need to separate substitution and income effects.

Consumer Theory	Producer Theory
Utility function $u(x)$	Production function $f(\mathbf{x})$
Consumption set $X \subseteq \mathbb{R}_+^L$	Production set $Y \subseteq \mathbb{R}^L$
Budget constraint $p \cdot x \leq w$	No budget constraint
Primal problems	
UMP: $\max u(x)$ s.t. $p \cdot x \leq w$	PMP: $\max p \cdot y$ s.t. $y \in Y$
EMP: $\min p \cdot x$ s.t. $u(x) \geq \bar{u}$	CMP: $\min w \cdot \mathbf{x}$ s.t. $f(\mathbf{x}) \geq q$
Value functions	
Indirect utility $v(p, w)$	Profit function $\pi(p)$
Decreasing in p , increasing in w	Increasing in output prices
Quasiconvex in p	Convex in p
Homog. degree 0 in (p, w)	Homog. degree 1 in p
Expenditure function $e(p, \bar{u})$	Cost function $c(w, y)$
Increasing, concave in p	Increasing, concave in w
Homog. degree 1 in p	Homog. degree 1 in w
Demand/supply	
Marshallian demand $x(p, w)$	Net supply $y(p)$
Hicksian demand $h(p, \bar{u})$	Conditional factor demand $\mathbf{z}(w, y)$
Envelope results	
Roy's identity: $x_i = -\frac{\partial v / \partial p_i}{\partial v / \partial w}$	Hotelling's lemma: $y_i = \frac{\partial \pi}{\partial p_i}$
Shephard (expenditure): $h_i = \frac{\partial e}{\partial p_i}$	Shephard (cost): $z_i = \frac{\partial c}{\partial w_i}$
Substitution effects	
Slutsky matrix $S = D_p h(p, \bar{u})$	Factor substitution $D_w \mathbf{z}(w, y)$
Symmetric, NSD	Symmetric, NSD
Net supply substitution	$D_p y(p) = D^2 \pi(p)$
Slutsky equation needed	Symmetric, PSD (no correction)
Laws	
Law of <i>compensated</i> demand:	Law of supply:
$\Delta p \cdot \Delta h \leq 0$	$\Delta p \cdot \Delta y \geq 0$
(requires Slutsky compensation)	(holds unconditionally)

7 Summary of Key Results

Production theory—key takeaways

1. A competitive firm sets $p = MC(y)$ on the upward-sloping portion of the marginal cost curve above AVC .
2. The supply curve is the inverse of the MC curve above the close-down price p^* .
3. Cost minimization requires $MRTS_{LK} = w/r$: the technical rate of substitution equals the factor price ratio.
4. The cost function $c(w, y)$ is concave in w , Homog. of degree one in w , and convex in q when f is concave.
5. Shephard's lemma: $\nabla_w c(\bar{w}, y) = z(\bar{w}, y)$ recovers conditional factor demands.
6. Hotelling's lemma: $\nabla_p \pi(\bar{p}) = y(\bar{p})$ recovers the supply function.
7. The law of supply $(p^1 - p^0) \cdot (y^1 - y^0) \geq 0$ holds without any compensation.
8. WAPM and WACM are nonparametric, assumption-free tests of profit-maximizing and cost-minimizing behavior, applicable to discrete data.
9. The properties of the cost function (homogeneity, concavity, monotonicity) are *sufficient* as well as necessary: any function with these properties is the cost function of some technology.

References

- Mas-Colell, A., M.D. Whinston, and J.R. Green (1995). *Microeconomic Theory*. Oxford University Press. Chapter 5.
- Varian, H.R. (1992). *Microeconomic Analysis*, 3rd ed. W.W. Norton. Chapters 1–6.
- McFadden, D. (1978). “Cost, Revenue, and Profit Functions.” In *Production Economics: A Dual Approach to Theory and Applications*, ed. M. Fuss and D. McFadden.
- Shephard, R.W. (1953). *Cost and Production Functions*. Princeton University Press.
- Diewert, W.E. (1974). “Applications of Duality Theory.” In *Frontiers of Quantitative Economics*, Vol. II, ed. M.D. Intriligator and D.A. Kendrick.

A Duality

In the preceding sections, we moved from the production function to cost and profit functions. A remarkable fact is that *this process can be reversed*: given a cost function, we can reconstruct the technology. This means the cost function contains essentially the same information as the production function. This general observation is the *principle of duality*.

A.1 Recovering the Technology from the Cost Function

Definition A.1 (Dual input requirement set). Given a cost function $c(w, y)$, define

$$V^*(y) = \{\mathbf{x} \geq 0 : w \cdot \mathbf{x} \geq c(w, y) \text{ for all } w \geq 0\}.$$

This is the set of all input bundles that cost at least as much as the minimum cost for every possible price vector.

Theorem A.1 (Fundamental duality theorem).

- (i) $V^*(y) \supseteq V(y)$ always.
- (ii) If the original input requirement set $V(y)$ is convex and monotonic, then $V^*(y) = V(y)$.
- (iii) If $V(y)$ is non-convex, then $V^*(y)$ is the convexification of $V(y)$, and $c^*(w, y) = c(w, y)$ (the cost function is the same for both).

Proof of (i). If $\mathbf{x} \in V(y)$, then $\mathbf{x}$ can produce q units, so $w \cdot \mathbf{x} \geq c(w, y)$ for all w (since c is the minimum cost). Hence $\mathbf{x} \in V^*(y)$. $\square$

Proof sketch of (ii). Suppose $\mathbf{x} \notin V(y)$. Since $V(y)$ is closed and convex, the separating hyperplane theorem gives a $w^* \geq 0$ such that $w^* \cdot \mathbf{x} < w^* \cdot \mathbf{z}$ for all $\mathbf{z} \in V(y)$. In particular, $w^* \cdot \mathbf{x} < c(w^*, y)$, so $\mathbf{x} \notin V^*(y)$. $\square$

The fundamental principle of duality

The cost function of a firm summarizes *all economically relevant information* about its technology. Two technologies with the same cost function are indistinguishable from the standpoint of economic behavior. If a technology is non-convex, its cost function is identical to that of its convexification—so convexity is essentially “free” from an economic standpoint.

A.2 Sufficient Conditions for Cost Functions

We know that all cost functions are Increasing, Homog. of degree 1, concave, and continuous in factor prices. Are these conditions also *sufficient*—does any function with these properties arise as the cost function of some technology?

Theorem A.2 (Characterization of cost functions). Let $\phi(w, y)$ be a differentiable function satisfying:

- (i) $\phi(tw, y) = t\phi(w, y)$ for all $t \geq 0$ (Homog. of degree 1 in w),
- (ii) $\phi(w, y) \geq 0$ for $w \geq 0$ and $q \geq 0$ (nonnegativity),
- (iii) ϕ is Increasing in each w_i (monotonicity),
- (iv) ϕ is concave in w (concavity).

Then ϕ is the cost function for the technology $V^*(y) = \{\mathbf{x} \geq 0 : w \cdot \mathbf{x} \geq \phi(w, y) \text{ for all } w \geq 0\}$.

Remark A.1. This is a powerful result for applied work: to specify a technology, it suffices to write down any function satisfying (i)–(iv). This may be much more convenient than specifying a production function directly, especially for econometric estimation, because the arguments of the cost function (factor prices) are typically exogenous.

A.3 Recovering Technology: An Example

Example A.1 (Cobb–Douglas recovery). Suppose we are given the cost function $c(w_1, w_2, y) = q \cdot w_1^\alpha w_2^{1-\alpha}$. By Shephard's lemma:

$$x_1(w_1, w_2, y) = \frac{\partial c}{\partial w_1} = \alpha q \cdot w_1^{\alpha-1} w_2^{1-\alpha}, \quad x_2(w_1, w_2, y) = (1 - \alpha) q \cdot w_1^\alpha w_2^{-\alpha}.$$

From the first equation: $w_2/w_1 = \frac{\alpha x_2}{(1-\alpha)x_1}$, so $w_1^\alpha w_2^{1-\alpha}$ can be expressed in terms of x_1, x_2 . Eliminating prices yields:

$$q = f(x_1, x_2) = A \cdot x_1^\alpha x_2^{1-\alpha}$$

for a constant A —the Cobb–Douglas production function. Thus the cost function and the production function carry the same information.

A.4 Completeness of Restrictions on Factor Demands

Proposition A.1 (Completeness). *Suppose functions $g_i(w, y)$, $i = 1, \dots, n$, satisfy:*

- (i) *Homogeneity of degree 0 in w : $g_i(tw, y) = g_i(w, y)$ for all $t > 0$.*
- (ii) *The substitution matrix $\partial g_i / \partial w_j$ is symmetric and negative semidefinite.*

Then there exists a technology whose conditional factor demands are exactly $\{g_i\}$.

Proof sketch. Construct $\phi(w, y) = \sum_i w_i g_i(w, y)$. Verify that ϕ satisfies the sufficient conditions of Theorem A.2: homogeneity of degree 1 follows from the degree 0 of g_i ; monotonicity and non-negativity follow from $g_i \geq 0$; concavity follows from the NSD substitution matrix. Then apply Theorem A.2. $\square$

Remark A.2. This means that homogeneity and negative semidefiniteness are the *complete list* of restrictions that cost-minimizing behavior places on conditional factor demands. Any functions satisfying these are factor demands for some technology. The parallel in consumer theory is that homogeneity, symmetry, negative semidefiniteness, and the budget constraint exhaust the restrictions on Hicksian demands.

A.5 Geometry of Duality

The duality between the production function and the cost function has a beautiful geometric manifestation.

Proposition A.2 (Curvature inversion). *The curvature of the isoquant and the curvature of the isocost curve³ are inversely related:*

- (i) *Highly curved isoquant (inputs are poor substitutes) $\longleftrightarrow$ nearly flat isocost curve (cost is insensitive to relative prices).*
- (ii) *Nearly flat isoquant (inputs are good substitutes) $\longleftrightarrow$ highly curved isocost curve (cost is sensitive to relative prices).*

In the extreme cases:

- *Leontief technology (L-shaped isoquant) $\longleftrightarrow$ linear cost function.*
- *Linear technology (flat isoquant) $\longleftrightarrow$ kinked (L-shaped) cost function.*

³The isocost curve is the locus of factor prices (w_1, w_2) that yield a given cost level: $\{(w_1, w_2) : c(w_1, w_2, y) = \bar{c}\}$.

Why curvature inverts

If the isoquant is sharply curved (nearly Leontief), then changing factor prices barely changes the optimal input mix. So the cost-minimizing bundle barely moves as prices change, meaning $\partial c / \partial w_i \approx \text{const}$ —the cost function is nearly linear. Conversely, if the isoquant is flat (nearly linear), a small price change causes a large shift in the optimal input—the cost function responds sharply, creating a kink.

B Efficient Production and Welfare

Definition B.1 (Efficient production). A production vector $y \in Y$ is *efficient* if there is no $y' \in Y$ such that $y' \geq y$ and $y' \neq y$.

Theorem B.1 (First fundamental theorem of welfare economics—production). *If $y \in Y$ is profit maximizing for some $p \gg 0$, then y is efficient.*

Proof. Suppose $y' \in Y$ with $y' \geq y$ and $y' \neq y$. Since $p \gg 0$, we have $p \cdot y' > p \cdot y$, contradicting the profit-maximality of y . $\square$

Theorem B.2 (Second fundamental theorem—production). *Suppose Y is convex. Then every efficient $y \in Y$ is profit maximizing for some nonzero $p \geq 0$.*

Proof. Define $P_y = \{y' \in \mathbb{R}^L : y' \gg y\}$. Since y is efficient, $Y \cap P_y = \emptyset$. By the separating hyperplane theorem (using convexity of Y and openness/convexity of P_y), there exists $p \neq 0$ such that $p \cdot y' \geq p \cdot y''$ for all $y' \in P_y$ and $y'' \in Y$. Since elements of P_y can be made arbitrarily close to y from above, $p \cdot y \geq p \cdot y''$ for all $y'' \in Y$, so y is profit maximizing at p . The nonnegativity $p \geq 0$ follows because $p_\ell < 0$ for some ℓ would allow unbounded profits via free disposal. $\square$

C Aggregation

Proposition C.1 (Aggregate supply). *Suppose there are J firms with production sets $Y_1, \dots, Y_J$. Let $\pi_j(p)$ and $y_j(p)$ be the profit function and supply correspondence of firm j . Define the aggregate production set $Y = Y_1 + \dots + Y_J$. Then for all $p \gg 0$:*

$$(i) \quad \pi^*(p) = \sum_{j=1}^J \pi_j(p).$$

$$(ii) \quad y^*(p) = \sum_{j=1}^J y_j(p).$$

Decentralization on the supply side

This is a decentralization result: to find the aggregate profit-maximizing production plan, it suffices for each firm to maximize profits *independently* at the common price p . No coordination is needed. Moreover, the aggregate supply matrix $D_p y(p)$ inherits symmetry and positive semidefiniteness, so the law of supply holds in the aggregate.

Remark C.1 (Convexification via aggregation). Even if individual production sets are nonconvex, the *average* production set $(1/J)(Y_1 + \dots + Y_J)$ becomes approximately convex for large J (Shapley–Folkman theorem). This is a regularizing effect analogous to aggregation in demand theory.

C.1 Necessary and Sufficient Conditions for Profit-Maximizing Behavior

We have seen that the cost function properties (homogeneity, concavity, monotonicity) are both necessary and sufficient for a function to be a cost function of some technology (Theorem A.2), and that homogeneity of degree 0 and symmetric NSD substitution matrix are the complete restrictions on conditional factor demands (Proposition A.1). We now ask the analogous questions on the profit-maximization side:

- (a) What conditions on a function $\pi(p)$ are necessary and sufficient for it to be the profit function of some technology?
- (b) What conditions on net supply functions $y(p)$ are necessary and sufficient for them to arise from profit maximization?
- (c) What conditions on discrete data $\{(p^t, y^t)\}_{t=1}^T$ are necessary and sufficient for them to be consistent with profit maximization?

C.1.1 Characterization of Profit Functions

Theorem C.1 (Necessary and sufficient conditions for profit functions). *A function $\pi : \mathbb{R}_{++}^L \rightarrow \mathbb{R}$ is the profit function of some closed, convex, free-disposal production set Y if and only if:*

- (i) $\pi(tp) = t\pi(p)$ for all $t > 0$ (homogeneous of degree one),
- (ii) π is convex,
- (iii) $\pi(p) \geq 0$ for all $p \gg 0$ (nonnegativity, reflecting $0 \in Y$).

The production set is recovered by

$$Y = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \text{ for all } p \gg 0\}.$$

Proof. Necessity. Properties (i)–(iii) were established in Proposition 5.3: (i) and (ii) hold for any closed Y with free disposal, and (iii) follows from $0 \in Y$ since $\pi(p) \geq p \cdot 0 = 0$.

Sufficiency. Given π satisfying (i)–(iii), define $Y = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \forall p \gg 0\}$. We verify:

Y is closed and convex: Y is the intersection of half-spaces $\{y : p \cdot y \leq \pi(p)\}$, one for each $p \gg 0$. The intersection of any family of closed convex sets is closed and convex.

$0 \in Y$: For all $p \gg 0$, $p \cdot 0 = 0 \leq \pi(p)$ by (iii), so $0 \in Y$.

Free disposal: If $y \in Y$ and $y' \leq y$, then $p \cdot y' \leq p \cdot y \leq \pi(p)$ for all $p \gg 0$ (since $p \gg 0$), so $y' \in Y$.

π is the profit function of Y : Define $\hat{\pi}(p) = \sup_{y \in Y} p \cdot y$. By construction of Y , $p \cdot y \leq \pi(p)$ for all $y \in Y$, so $\hat{\pi}(p) \leq \pi(p)$. For the reverse, by the supporting hyperplane theorem for convex functions: since π is convex and homogeneous of degree one, for each $\bar{p}$ there exists a subgradient $\bar{y}$ with $\pi(p) \geq \pi(\bar{p}) + \bar{y} \cdot (p - \bar{p})$ for all p . By homogeneity of degree one, $\pi(\bar{p}) = \bar{y} \cdot \bar{p}$ (Euler's theorem), so $\pi(p) \geq \bar{y} \cdot p$ for all $p \gg 0$. Hence $\bar{y} \in Y$ and $\hat{\pi}(\bar{p}) \geq \bar{p} \cdot \bar{y} = \pi(\bar{p})$. Thus $\hat{\pi} = \pi$. $\square$

Remark C.2 (Parallel with cost functions). Compare with Theorem A.2: a cost function must be homogeneous of degree one, concave, and nondecreasing in factor prices. A profit function must be homogeneous of degree one, convex, and nonnegative. The cost function is concave because cost minimizers re-optimize to *reduce* cost when prices change; the profit function is convex because profit maximizers re-optimize to *increase* profit when prices change.

C.1.2 Characterization of Net Supply Functions

Theorem C.2 (Necessary and sufficient conditions for net supply functions). *Functions $y_\ell(p)$, $\ell = 1, \dots, L$, are the net supply functions of some technology if and only if:*

- (i) $y(tp) = y(p)$ for all $t > 0$ (homogeneous of degree zero),
- (ii) The substitution matrix $S(p) = D_p y(p)$ is symmetric: $\frac{\partial y_\ell}{\partial p_k} = \frac{\partial y_k}{\partial p_\ell}$ for all ℓ, k ,
- (iii) $S(p)$ is positive semidefinite: $v^\top S(p) v \geq 0$ for all $v \in \mathbb{R}^L$.

Proof. Necessity. If $y(p) = \nabla \pi(p)$ for a profit function π (Hotelling's lemma), then:

- (i) follows from π being HD1: differentiating $\pi(tp) = t\pi(p)$ with respect to p gives $t\nabla \pi(tp) = t\nabla \pi(p)$, so $\nabla \pi(tp) = \nabla \pi(p)$, i.e., $y(tp) = y(p)$.
- (ii) follows from Young's theorem: $S(p) = D^2 \pi(p)$ is symmetric.
- (iii) follows from convexity of π : the Hessian of a convex function is PSD.

Sufficiency. Given functions $y(p)$ satisfying (i)–(iii), we construct a profit function π whose gradient is y , then apply Theorem C.1.

Step 1: Integrability. Conditions (i) and (ii) are exactly the integrability conditions ensuring that $y(p)$ is the gradient of some function. Symmetry of $S(p) = D_p y(p)$ means the vector field y is conservative, so there exists $\pi : \mathbb{R}_{++}^L \rightarrow \mathbb{R}$ with $\nabla \pi(p) = y(p)$. Concretely, fix any reference price $\bar{p} \gg 0$ and define

$$\pi(p) = \int_0^1 y(\bar{p} + t(p - \bar{p})) \cdot (p - \bar{p}) dt + \pi(\bar{p}),$$

where $\pi(\bar{p})$ is an arbitrary constant (say $\pi(\bar{p}) = \bar{p} \cdot y(\bar{p})$, which is the natural choice by Euler's theorem). The symmetry of $D_p y$ guarantees that this line integral is path-independent, so π is well-defined.

Step 2: π is HD1. By homogeneity of degree zero of y :

$$\nabla \pi(tp) = y(tp) = y(p) = \nabla \pi(p).$$

Since π and $t\pi$ have the same gradient (using $\nabla[t\pi(p)] = t\nabla \pi(p)$, and differentiating $\pi(tp)$ gives $t\nabla \pi(tp) = ty(p)$), the function $\pi(tp)/t$ has the same gradient as $\pi(p)$, so they differ by a constant. Evaluating at any point and using the choice $\pi(\bar{p}) = \bar{p} \cdot y(\bar{p})$ confirms $\pi(tp) = t\pi(p)$.

Step 3: π is convex. Since $D^2 \pi(p) = D_p y(p) = S(p)$ is PSD by condition (iii), π is convex.

Step 4: π is nonneg. By HD1 and Euler's theorem, $\pi(p) = p \cdot \nabla \pi(p) = p \cdot y(p)$, which can be set to be nonneg by appropriate normalization (or follows if $0 \in Y$).

By Theorem C.1, π is the profit function of some technology Y , and by construction $y(p) = \nabla \pi(p)$ is the net supply function. $\square$

Completeness on the supply side

Homogeneity of degree zero, symmetry, and positive semidefiniteness are the *complete list* of restrictions that profit-maximizing behavior places on net supply functions. Any functions satisfying these three conditions arise as net supply functions of some technology.

The parallel on the cost side (Proposition A.1) is: homogeneity of degree zero and symmetric NSD substitution matrix are the complete restrictions on conditional factor demands.

	Net supply $y(p)$	Conditional factor demand $z(w, q)$
Homogeneity	HD0 in p	HD0 in w
Symmetry	$\partial y_\ell / \partial p_k = \partial y_k / \partial p_\ell$	$\partial z_i / \partial w_j = \partial z_j / \partial w_i$
Definiteness	PSD	NSD
Source	$D^2\pi$ (convex)	$D^2_{ww}c$ (concave)

The sign difference (PSD vs. NSD) reflects the fundamental asymmetry: profit is *maximized* (the profit function is convex), while cost is *minimized* (the cost function is concave in factor prices).

C.1.3 Characterization from Discrete Data (WAPM)

When the economist observes not smooth functions but a finite set of price-quantity pairs, the characterization takes a different form.

Theorem C.3 (WAPM is necessary and sufficient). *A finite dataset $\{(p^t, y^t)\}_{t=1}^T$ of prices and production plans is consistent with profit maximization by some production set Y (i.e., there exists a closed, convex Y such that $y^t \in \arg \max_{y \in Y} p^t \cdot y$ for all t) if and only if WAPM holds:*

$$p^t \cdot y^t \geq p^t \cdot y^s \quad \text{for all } t, s = 1, \dots, T.$$

Proof. Necessity. If y^t maximizes $p^t \cdot y$ over Y , and $y^s \in Y$, then $p^t \cdot y^t \geq p^t \cdot y^s$ by definition of maximization.

Sufficiency. Assume WAPM holds. Construct:

$$Y_O = \{y \in \mathbb{R}^L : p^t \cdot y \leq p^t \cdot y^t \text{ for all } t = 1, \dots, T\}.$$

This is a closed convex set (intersection of half-spaces).

$y^t \in Y_O$ for all t : WAPM gives $p^s \cdot y^t \leq p^s \cdot y^s$ for all s , so y^t satisfies every inequality.

y^t maximizes $p^t \cdot y$ over Y_O : For any $y \in Y_O$, the t -th inequality gives $p^t \cdot y \leq p^t \cdot y^t$. Hence $y^t \in \arg \max_{y \in Y_O} p^t \cdot y$. $\square$

Summary: three levels of characterization

Profit-maximizing behavior can be characterized at three levels of generality:

- (i) **Discrete data** $\{(p^t, y^t)\}$: WAPM ($p^t \cdot y^t \geq p^t \cdot y^s$) is necessary and sufficient. No differentiability or functional-form assumptions needed.
- (ii) **Supply functions** $y(p)$: Homogeneity of degree 0, symmetric and PSD Jacobian are necessary and sufficient. This is the “integrability” characterization.
- (iii) **Profit function** $\pi(p)$: Homogeneity of degree 1, convexity, and nonnegativity are necessary and sufficient. The technology is recovered via $Y = \{y : p \cdot y \leq \pi(p) \forall p\}$.

Each level can be derived from the others: integrating supply functions gives the profit

function; differentiating the profit function gives supply; evaluating at finitely many prices gives discrete data. The characterizations are mutually consistent.