Lectures on Neoclassical Production Economics

Table of Contents

Lectures on Neoclassical Production Economics[1]

Lecture 0: Introduction

Robert G Chambers[2]

January 9, 2024

[2]University of Maryland

Students in economics are introduced to broad concepts in early ECON classes and expected to master them for more advanced study. Two such concepts are *supply* and *demand*. Suppliers are individuals who sell goods or commodities in market exchanges, and demanders are the buyers. These two forces are illustrated using a "scissors diagram" with a downward sloping demand curve and an upward sloping supply curve.

These lectures are a guide to the fundamental economic model behind the supply shear of the demand and supply scissors. Known as the "the theory of the producer" or "production economics", it describes the economic behavior of individuals who "make" commodities to sell in a market. "Make" is interpreted to mean both physical production of a commodity (for example, growing wheat) and delivering a service (for example, providing a taxi ride).

These lectures emerged from an undergraduate production economics class at the University of Maryland. In them, I try to distill the essence of that theory in a form that is accessible to anyone with a rudimentary class in microeconomics and the usual doses of high-school and first-year college mathematics. The essential ideas are due to others, and no claim is made to originality. I have tried, however, to make the lectures rigorous but accessible. Calculus-based arguments have been avoided. Many years of experience have taught me that grounding economic concepts in calculus terms too often promotes mechanical thinking that inhibits reasoning about more nuanced economic behavior. So while the calculus makes cameo appearances, it never assumes a supporting, let alone a starring, role. Instead, the argument relies on a geometric approach, grounded in convex analysis, that is at the same time more basic but more rigorous. The key ideas are conveyed with pictures, but they are rooted in results that I trace to R. T. Rockafellar's *Convex Analysis* and J.J. Moreau's *Fonctionnelles Convexes* that form the basis of much of modern optimization theory. In no small way, much of what I write here repeats arguments and analysis made in my book *Competitive Agents in Certain and Uncertain Market*, which was intended for a graduate-student audience, in a form that I hope is digestible to advanced undergraduate students in economics.

Another characteristic of these lectures is that they take a "constructive" approach to production economics. They start with simple concepts that reflect common axioms in modern production theory. Then they construct a representation of a technology by postulating the existence of observations on inputs and outputs and imposing those axioms on the data.

The lectures start with a general discussion of economic models. The takeaway message is that economic models do not depict reality. Instead, they are intellectual tools, organized around the principle of abstraction, that assist thinking about reality. Models have realistic kernels, but they contain unrealistic elements. Three chapters on different ways that economists model the production of goods and services follow. I first introduce and discuss the notion of the "technology". And then starting from simple cases, I consider more general structures. Four lectures on the neoclassical producer model follow. Each examines producer behavior from a different perspective. But all assume, without apology, that producers are profit seekers operating in a competitive environment. That means self interest motivates their behavior and that they cannot affect the prices at which they buy or sell. Lecture 9 shows that the path taken in the preceding four lectures is reversible. If producer behavior is consistent with accepted economic principles, there must exist a technology underpinning it that is consistent with the technology model developed in Lectures 2 through 4. So, in a way, Lecture 9 justifies Lectures 2-4's focus on the model of the technology.

Lectures on Neoclassical Production Economics[1]

Lecture 1: Models in Economics

Robert G. Chambers[2]

January 9, 2024

[2]University of Maryland

The Role of Models in Economics

> ...reality may avoid the obligation to be of interest...hypotheses may not..
>
> Lönrott to Treviranus in *Death and the Compass* by Jorge Borges

Parables are stories meant to convey a *moral*. They are simple so that the listener can grasp the moral easily. Good economic *models* are similar. They identify a problem and strip it of unnecessary detail to focus on its essence. Thus, economic models are *abstract* rather than realistic. Abstraction promotes analysis, but it also makes economic models *as if* in nature. The principle isn't that models are realistic but that valuable lessons are learnt by pretending *as if* they are.

As children, we learn something about transportation by playing with toy cars and something about cooking by playing with toy kitchens. Both a toy car and a toy kitchen are models. Both convey information about their realistic counterparts. And just as children enjoy playing with toys, they learn from them. Economists try to design their models so that they, too, learn real-world lessons from them.

Economic models assume verbal, mathematical, and visual (read graphical) forms. Often, they mix all three, and the forms are *always* interchangeable. Mathematical arguments translate into plain English. If they don't, something's wrong. By the same token, a verbal economic argument should be expressible mathematically. Economists illustrate their verbal and mathematical arguments using two or three-dimensional graphs. Examples with which you may be familiar include *demand* and *supply curves*. Just as mathematical arguments translate into words, these pictures translate into either words or mathematics.

Economists use mathematics and graphs as tools to promote precision, to ensure logical reasoning, to promote intuition, and to facilitate generalization from simpler to more complex settings. Mathematical symbols also provide a 'shorthand' that replaces (harder to write) strings of words. Once a symbol is defined, it can be used with other symbols to construct meaningful (and precise) sentences. Graphs illustrate concepts phrased in either verbal or mathematical terms in the belief that the adage, "a picture is worth a thousand words", is correct.

In working with a model, it's crucial to stay within its confines. Results must trace to an element or elements of your model. They cannot emerge from factors implicitly imported into your reasoning. If your model cannot capture behavior you wish to explain, consider changing or adapting it to incorporate elements that might explain the phenomenon.

Thought Experiments and Models

Albert Einstein often illustrated complex reasoning using simple *thought experiments*. For example, in *Relativity: The Special and General Theory*,[1], he used images of a moving railway carriage, a man standing on the railway carriage, a man on a railway platform, and lightning strikes to explain his *Special Theory of Relativity*. His thought experiments are examples of a rhetorical method that uses models spelt out in imaginary settings to convey an argument.

Thought experiments are important in economics, and most conceptual models are designed to be used with them. Typically, they are hypothetical. The reader is asked to imagine a plausible situation and then to make inferences about how individuals behave in that setting. Their importance in economics emerges from the difficulty that economists have in conducting *controlled experiments*. From your high-school science classes, you know that a *controlled experiment* is one in which all possible sources of variation for the outcome except one (the treatment) are held fixed.

Thinking about the complexity of the settings in which social transactions occur reveals the difficulty of achieving such control in studying them. Physicians, sociologists, psychologists, neuroscientists, and economists strive to execute such experiments, but the outcomes are often problematic. And with good reason. Imagine, for example, how two balls moving towards one another react as they strike. A bit of physics gives the answer. Now imagine the same setting when the two balls can think and control their trajectory. What happens then? Social scientists study problems that are similar to those physicists study in that they can involve physical interactions, but they differ in that those interactions involve thinking individuals. True control becomes difficult if not impossible.

[1] Albert Einstein. 1920. *Relativity: The Special and General Theory*. Methuen and Co. Ltd. Translated by Robert W. Lawson. Part 1, Section 09. Einstein intended this book for the general reader.

Some Simple Mathematical Background

...plus nous avançons dans la connaissance de l'univers, plus il nous apparaît fondè sur des lois mathèmatiques.[2]

Hervè Le Tellier, *L'anomalie.* Gallimard.

Many readers may find repeated mentions of 'mathematics' daunting. *Math anxiety*[3] is a common affliction that haunts many from elementary school on. Having suffered it, I can empathize. But if you are afflicted, please take a deep breath and read on. For while these lectures contain many mathematical symbols, that doesn't mean that the underlying ideas that the symbols represent are unfamiliar or something to which you cannot attach an everday meaning. All can and will be explained in verbal and pictorial terms. While we strive to maintain formal correctness, the focus is on the intuition that lies behind the symbols and not on the symbols themselves. So where most economics texts for advanced undergraduates freely use the calculus, these lectures avoid it precisely because so many find it impenetrable and unintuitive.[4] Still, math statements percolate throughout each of the lectures. My reasons for insisting on using mathematics are two: I believe in the sentiment expressed in this section's epigraph. And, much of what needs to be said requires some math notation and arguments.

Some, perhaps many, may find the notion that human behavior can be depicted using mathematics repellent because it conflicts with the belief that humans exercise free will. Such arguments are to be taken seriously, but a detailed discussion of them would turn what is intended as an introduction to an area of economics into a philosophy tome. But if you

[2]...the more we understand the laws of the universe, the more it seems that they are based on the laws of mathematics (my translation)

[3]This is particularly true for readers from the United States, where all-too-many believe that math is only accessible to the Sheldon Coopers of the world and not to regular people. And because so many feel this way, innumeracy is tolerated to an extent that illiteracy is not. Often it's rooted in experiences with teachers who, in addition to being sufferers themselves, project their own discomfort with rudimentary concepts into feelings of inadequacy upon the part of the student. On the other hand, there are those who simply suffer math antipathy, which is perfectly understandable.

[4]I also believe it promotes mechanical thinking about economics that tends to retard rather than promote a deeper intuitive understanding.

are so inclined, I would ask that you consider several points.

The physical reality in which we operate is described by mathematical principles to an astonishing degree of accuracy. Many things we now understand to be true about it were discovered as a consequence of mathematical argument. Why would inhabitants of that reality be exempt from such description? Saying that human behavior is amenable to mathematical representation is not the same as saying humans are mindless automatons. Indeed, the field of mathematics known as *optimization* is devoted to finding the best choice from an array of alternatives. And, more prosaically, how does one can explain humanity's obvious affinity for straight lines? We see them everywhere from reforested land to footpaths between two objects despite Kent's famous admonition that "Nature abhors a straight line". You need not be Hari Seldon to believe in the usefulness of mathematics in the study of human behavior.[5]

Regardless of how you feel about such matters, these Lectures are about economics and not mathematics. You should be able understand the fundamental story by skipping the math and just reading the words. However, that's akin to the golfer who avoids certain clubs. It gets you through the round and may even lower your score. But you'll still have the weaknesses in future rounds. Another, which I use, is to skim the equations that appear opaque on first reading, but revisit them once you have a grasp of the argument.

Because mathematical concepts and symbols are used, you need to know what they mean when you encounter them. Therefore, we now introduce a few concepts and a list of symbols. The intent of the latter is to provide a convenient reference for later discussion.

The most basic concept is a *set*, which is defined as *a collection of objects*. What those objects are can vary. Familiar examples include a set of tools, a set of silverware, or a dozen apples. Each meets the criterion of having objects in it. Perhaps more complicated, but still familiar, examples would be the set of integers and the set of rational numbers.[6] Objects in the set are referred to interchangeably as *members* or *elements* of the set.

We denote sets by the symbol $\{\cdot\}$. Here, and in all that follows, the symbol $\cdot$ is a stand-in

[5]William Kent was an 18th century English landscape architect. Hari Seldon is a fictional psychometrician who is the hero of Issac Asimov's *Foundation*.

[6]Integers are the whole numbers such as 1,2,3 and a rational number is one that can be expressed as a ratio of two integers.

for an argument to be filled in later. The set definitions we use in these lectures have three parts *the name of the set*, *a description of the kind of objects that are in the set*, and *a qualifying statement about those objects*. For example, the expression

$$O = \{\text{oranges: the oranges are edible}\},$$

is to be read as "O is the set of oranges such that (a colon, :, is the math symbol for the words *such that*)[7] the oranges are edible". Here the name of the set is O, the kind of objects in the set are oranges, and the qualifier is that the oranges be edible. Similarly,

$$RO = \{\text{oranges: the oranges are inedible}\}$$

would be be read as RO is the set of oranges that are inedible.

If all the elements of set A also belong to set B, then A is a *subset* of B, which we denote by $A \subset B$. For example, if $B = \{\textit{ all apples}\}$ and $A = \{\textit{golden apples}\}$ then $A \subset B$. The *intersection* of sets A and B, denoted $A \cap B$, consists of the set of elements that they have in common. For example, if $A = \{\textit{red}, \textit{white}, \textit{blue}\}$ and $B = \{\textit{white}, \textit{black}, \textit{yellow}\}$ then $A \cap B = \{\textit{white}\}$. Unless assumed otherwise, the order in which elements of a set appear is arbitrary. Therefore, $A = \{\textit{red}, \textit{white}, \textit{blue}\}$ and $B = \{\textit{red}, \textit{blue}, \textit{white}\}$ are the same set, denoted $A = B$. Alternatively, $A \subset B$ and $B \subset A$ implies that $A = B$.

Associated with the notion of a set is that of the *empty set*, which is defined as the set with nothing in it. You might reasonably ask why we are interested in a set that has nothing in it? My best answer is that we often encounter situations that involve physical or (currently) technical impossibilities. And to tell complete economic stories, we need to account for them. The notion of an empty set lets us. The notation for the empty set is $\emptyset$, which resembles a zero with a backslash drawn through it. The convention is that $\emptyset$ belongs to all sets.

Common Symbols We now list, for convenient later reference, some symbols used in these lectures. We haven't yet told you what they all mean, but that will come later.

- $\{\cdot\}$, set

[7] Many writers use $|$, vertical slash, in place of :.

- $\emptyset$, the empty set
- $\subset$, is a subset of, $A \subset B$ reads A is a subset of B
- $\cap$, intersection, $A \cap B$ is the intersection of sets A and B
- x, a single input or a vector of inputs
- w, a single input price or a vector of input prices
- y, a single output or a vector of outputs
- p, a single output price or a vector of output prices
- $wx = \sum_n w_n x_n$, the sum of all input prices times the input values
- $py = \sum_m p_m y_m$, the sum of all output prices times the output values
- $\exists$, there exists
- :, such that
- $\geq$, greater than or equal to
- $\leq$, less than or equal to
- $=$, equal to or "*is*"
- $\equiv$, is defined as or is equivalent to
- (a, b), all real numbers x such that $a < x < b$
- $[a, b]$, all real numbers x such that $a \leq x \leq b$
- ΔA, the change in A
- $\Rightarrow$ *or* $\Downarrow$, implies. For example, $A \Rightarrow B$ is to be read as A implies B. Alternatively (and more completely), $A \Rightarrow B$ is translated as: If A is true, then B is true.
- $\Leftrightarrow$, two statements are equivalent. For example, $A \Leftrightarrow B$ is to be read as A and B are equivalent, or as A implies B and B implies A.

- for vectors x and y, $x \geq y, x \neq y$, every element of x is at least as large as the corresponding element in y and one element of x is strictly larger than its corresponding element in y. For example, $(0,0)$ and $(0,1)$ both satisfy $(0,0) \geq (0,0)$ and $(0,1) \geq (0,0)$ but only $(0,1) \geq (0,0), (0,1) \neq (0,0)$.

- $\in$, belongs to or is a member of. For example, $\hat{a} \in A$ is to be read as $\hat{a}$ belongs to A, or $\hat{a}$ is a member of A

- $\notin$, does not belong to or is not a member of. For example, $\hat{a} \notin A$ is to be read as $\hat{a}$ is not a member of A.

Lectures on Neoclassical Production Economics[1]

Lecture 2: The Producer and the Technology

Robert G. Chambers[2]

January 9, 2024

[2]University of Maryland

Producers

We study individuals who make (*produce*) goods or services to sell in markets. We refer to them interchangeably as *producers*, *decisionmakers*, or *firms*. A handy example is a farmer producing agricultural commodities (for example, wheat, corn, soybeans, hogs, beef, etc.). But feel free to substitute your own examples to suit your experience and intuition. The more that you think in familiar terms the more you will understand.

We make basic assumptions about what motivates producers and about the setting in which they make decisions. **These assumptions are maintained throughout the lectures.** You don't need to agree with them or even find them believable. But the analysis that follows assumes that they hold.

The first assumption is that producers are *small* or *perfectly competitive* relative to the markets that they face. Small means that they take the prices at which they can buy and sell market items as *given by the market.* Producers cannot manipulate prices, they are *price takers.* The second assumption is that producers are *self-interested* individuals who seek to maximize their profit from producing and selling commodities.

In other economics classes, you may have encountered situations where producers can manipulate prices (for example, monopoly or monopsony). These lectures aren't about them. We treat a situation similar to what one faces on entering a *Starbucks@* or *McDonalds@*. You do not bargain with a barista over the price of a cappucino. Either you pay the stated price or you don't get the cappucino. Same here, a producer doesn't negotiate the price of an input or the selling price of the product.

The profit maximizing assumption can be controversial. Many argue that individuals are not motivated solely by self gain. Others argue that producers should not behave selfishly. I make no argument either for or against those positions. *But* these lectures are about individuals who, when faced with a market setting, maximize profit. If producers maximize profit, what prevents them from making it infinitely large? Our answer is what we call the *Technology*, which is the subject of the rest of this Lecture.

The Technology Set Introduced

The *Technology* describes the possibilities for producing *outputs* using *inputs*. *Inputs*, which we denote by x, are commodities or goods used to produced something. For example, in making a cake the inputs include its raw ingredients, your time and effort, heat from the oven, etc. *Outputs*, denoted by y, are the commodities produced (for example, the cake). We characterize the *feasible* possibilities as a set that contains *inputs* and *outputs* and that is common to all producers. Producers cannot affect that set. So, it is not unlike market prices. Producers take it as given and then react to it in making their production decisions. You might think of this set as given by *Nature*.

The mathematical definition is

$$T = \{(x, y) : x \text{ can produce } y\}$$

Here T is a notational stand-in for the *Technology*. The expression (x, y) tells us that the members of the set are inputs and outputs, and the qualifying statement is that "x can produce y". Expressed in words: *The technology is the set of inputs and outputs such that the outputs can be produced using the inputs.*[1] So, saying that "x and y belong to T", written in symbols as $(x, y) \in T$, is the same as saying "x can produce y". Or, in symbols, $(x, y) \in T \Leftrightarrow x$ can produce y. (Recall that $\Leftrightarrow$ translates into words as "is equivalent to".)

Thus, the technology contains all feasible combinations of inputs and outputs. This way of thinking of the technology contrasts with every day usage of the word *technology*. For example, you may have heard individuals speak of the *technology for growing corn* or the *technology for producing steel*. This is sensible given most dictionary definitions of technology. But it's *not* how we use the term. Our usage of the term includes all such notions in a common set from which any producer can choose. So, instead of saying the *technology for growing corn* we use the phrasing the *process for growing corn* or *the technical process for growing corn*.

[1] Having introduced and defined the term, in the following we refer to the technology using lower-case letters.

The Technology is a "Model"

T contains many possibilities, most of which are complex. That's true for even simple choices. Think of what's involved in producing a wooden pencil. Its obvious inputs include wood, pencil lead, an eraser, and a piece of metal or plastic to hold the eraser in place. Each of these inputs have themselves been produced or processed. The productive process stretches further back. It quickly becomes apparent that keeping track of all such activities becomes complex. Faced with such complexity, we need a *model*. This chapter details the production model used in these lectures. We construct it by imposing assumptions (restrictions) upon T. We strive for a model that captures the economic essentials of production but that remains parsimonious.

First, we must clarify how we think about x and y. We have said that they are inputs and outputs but not how we measure them. We assume they can be represented as *real numbers*. Because we are interested in cases that may involve multiple inputs and multiple outputs we let x be a nonnegative N-dimensional vector

$$x = (x_1, x_2, ..., x_N),$$

where x_n stands for the *nth* input to the production process. Similarly, y stands for a nonnegative M-vector

$$y = (y_1, y_2, ..., y_M),$$

and y_m for the *mth* output produced. M and N are integers and typically differ. In the following, to avoid mathematical jargon, we often refer to $x's$ and $y's$ that are vectors as *bundles* and not as vectors. The concepts are interchangeable.

Real numbers are familiar, so it's easy to gloss over this assumption. But familiar or not, they are abstract and a bit mysterious. They represent the completion of the rational numbers. Or, put another way, they are the set of rational numbers augmented by the numbers (the irrational numbers) that fall in between the rational numbers to form the continuum. Mathematicians developed the notion of an irrational number because physical phenomena exist that we cannot measure exactly without them. *Pi* that gives the ratio of a circle's circumference to its diameter is one example. Another is the length of the diagonal of a square. By Pythagoras's Theorem the length of the diagonal for a square with 90 foot sides is $\sqrt{2} \times 90$ feet. And, regardless of the units used to measure the square's sides, the diagonal's length remains $\sqrt{2}$ times that length. That means that we cannot devise a discrete system of measuring the lengths of diagonals and the sides. It's as if Nature demands the existence of an irrational number, $\sqrt{2}$, to measure length.

In thinking about inputs and outputs, it's important to remember that T is defined in terms of *all* possible inputs and outputs. In particular, x contains all possible inputs to a production process and y contains all the possible outputs. To give an example, think of a farmer who grows corn on a parcel of land. If asked to list the associated inputs, one might respond land, labor, fertilizers and pesticides, capital equipment, and fuel. That list, in fact, encompasses the inputs most often considered in empirical studies of farmer behavior, but it is incomplete. Corn growth also depends upon other factors including natural moisture, natural radiation, the presence or absence of a pest infestation, the organic content of the soil, etc. In short, anything that affects corn growth is included in x. x even includes factors that inhibit production, such as pests.

What's true for x is true for y. When we write y, we mean all outputs. For the corn grower that extends the list beyond corn to include air pollution emitted by burning fuel, soil run-off as a result of plowing, fertilizer and pesticide leakage into the environment, noise pollution from running capital equipment, etc. Again the list contains both desirable outputs and nondesirables such as pollution, noise, and runoff. All belong in T.

We emphasize the higher dimensional nature of x and y because most of our discussions are in simpler terms. For example, our graphical illustrations are two (sometimes three) dimensional. So, when we discuss the interaction of inputs and outputs, it will be rare for us to illustrate a production process that contains more than one or two outputs and one or two inputs. That doesn't mean that other dimensions are not important. Instead it manifests our visual perception of the world as three-dimensional and reading surfaces as two-dimensional. In drawing such pictures, we assume that *all other things are held constant.*[2]

As you read, please keep in mind that our model of T assumes a deterministic setting.

[2] In economics jargon, the Latin phrase *ceteris paribus* is often used in place of "all other things held constant".

We do not treat situations where producers make decisions in the presence of uncertainty about the outcomes of those decisions. This is unrealistic, but we are trying to build a tractable model and not to replicate reality. And, as unrealistic as it is, it's a common assumption in even the most advanced economic analyses. No one believes that uncertainty is unimportant. But treating uncertainty makes even the simplest production problems complicated. We also assume that T satisfies a *regularity condition*:[3] T is a *closed set* and $\{y : (x, y) \in T\}$ is bounded for all x finite. If you have a smattering of math knowledge, you may know what this means. If you don't, the regularity condition's primary role is to simplify the definition of maxima and minima in what follows.

Assumptions on T

Our restrictions on T mix five assumptions. We list them in mathematical form and then explain each in the subsections that follow. A good way to view these assumptions is as a menu of alternative properties T may possess. Each assumption may be reasonable in some settings but not in others.

Assumptions on T

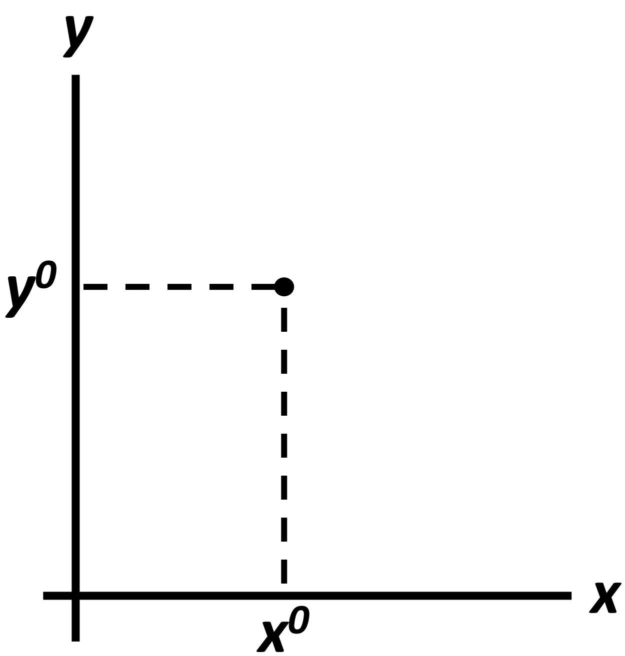
- 1. $T \neq \emptyset$ (T is nonempty);
- 2. $(x, y) \in T \Rightarrow (x', y) \in T$ for $x' \geq x$ (Free disposability of inputs (FDI));
- 3. $(x, y) \in T \Rightarrow (x, y') \in T$ for $y' \leq y$ (Free disposability of outputs (FDO));
- 4. If $(x^0, y^0) \in T$ and $(x^1, y^1) \in T$ then $\lambda (x^0, y^0) + (1 - \lambda) (x^1, y^1) \in T$ for all $\lambda \in (0, 1)$ (T is a *convex* set).
- 5. $(0, y) \notin T$ if $y \geq 0$ and $y \neq 0$ (No Free Lunch (NFL)) and $(0, 0) \in T$ (No Fixed Costs (NFC)).

[3] In a more formal setting, the regularity condition would be listed as a separate (the sixth) assumption. I have avoided this to maintain the focus on what I regard as the more important restrictions.

T is nonempty

$T \neq \emptyset$, which is read as "T is not an empty set", is easy to justify. It means that we assume that an x and a y exist such that x can produce that y. Because we live in a world where things are produced, it's plausible.

Figure 1: T not empty



Nonemptiness of T is a good place to start illustrating our model of T. In all illustrations that follow, we assume that you understand the traditional $x - y$ axis representation of the Cartesian coordinate system (coordinate plane). We refer to the $x - y$ axis representation interchangeably as (x, y) *space* and the (x, y) *plane*.

In Figure 1, we label the vertical axis as y and let it measure output. We use the horizontal axis to measure input and label it x. Because there are only two dimensions, the figure is a stylized representations of our ideas. Various interpretations exist. It can illustrate a hypothetical technology where one input produces a single output, how one output and one input covary while holding all others fixed, or, in a macroeconomic spirit, the behavior of an "aggregate input" and an "aggregate output". Nonemptiness of T means that at least one feasible point exists in (x, y) space. It's illustrated by the point (x^0, y^0).

Free disposability of inputs (FDI)

FDI requires that if x can produce y, any input *bundle* that is at least as large as x in all its N dimensions can produce that same y. This assumption strikes many as intuitive. And in many cases it is. For example, if a farmer can produce 75 bushels of wheat with one hired hand, it seems reasonable that the farmer could continue to produce 75 bushels of wheat if another hired hand were added. In fact, intuition would suggest that the farmer could produce more than 75 bushels. Intuition might also suggest that if the farmer increased the other inputs devoted to wheat production, producing 75 bushels of wheat remains feasible.

Starting from the feasible point, (x^0, y^0), depicted in Figure 1, FDI requires that all points to its right holding output at y^0 are feasible. Figure 2 illustrates. Note the logical force of the assumption. Nonemptiness of T ensures that one feasible input-output bundle exists. Adding FDI ensures an infinity of feasible input-output combinations exist.

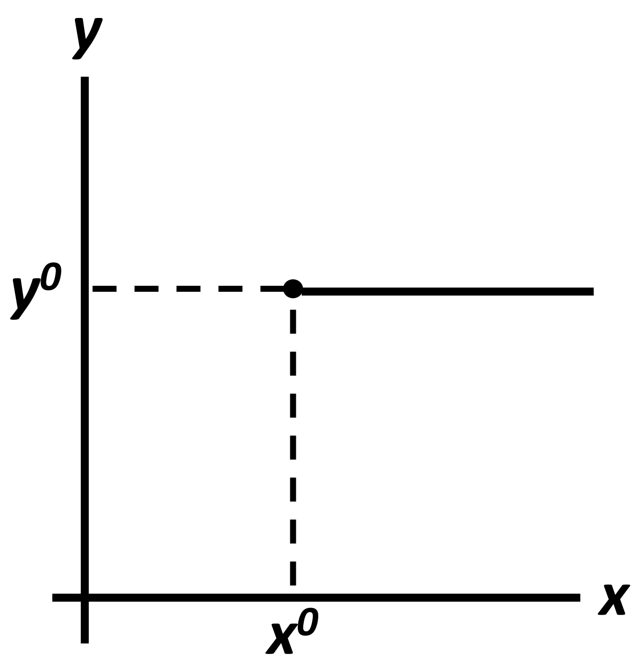
This intuition has its limits. To illustrate, let's use our farmer example. Suppose that the farmer's inputs are land, hired labor, fertilizer, and seed and that the output is wheat. Let's also say that the farmer's input usage to produce that 75 bushels is

$$x = (1, 1, 50, 40),$$

which stands for 1 unit of land, 1 unit of hired labor, 50 units of fertilizer, and 40 units of seed. (You are free to choose 'units' to suit your intuition.) Now, what do you believe happens if land, hired labor, and seed remain constant but fertilizer increases *without bound*? In Figure 2, if the horizontal axis measured fertilizer, we illustrate that by allowing x to run off the edge of the page and on to infinity.

FDI requires that producing 75 bushels of wheat remains feasible. But this counters reality. It is well known that applying *too much* fertilizer is destructive rather than productive.

Figure 2: Free disposability of input



Production economists call this phenomenon *congestion*. It happens because increasing one input while holding others constant eventually overwhelms the other inputs with which it cooperates. Another simple example comes from holding land, fertilizer, and seed constant at $(1, 50, 40)$, respectively, and allowing hired labor to grow towards infinity. Sooner rather than later, those extra workers get in one another's way. Congestion is a general phenomenon common to many different production processes. The aphorisms that one can "get too much of a good thing"or that "too many cooks can spoil the broth" capture the intuitive essence of congestion.

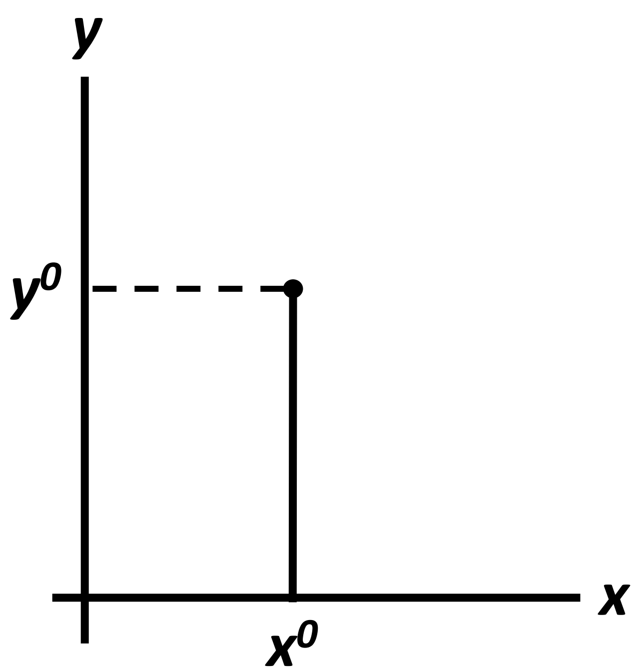
So, if FDI counters physical reality, why use it? The answer is that it is an acceptable approximation in many practical instances. Think of our farmer. Because the farmer knows the technology, the farmer will know about congestion and will not operate in regions of T where it exists. It's hard to imagine a producer applying destructive inputs. If they do, they soon go out of business. Therefore, while we know congestion can occur, we may lose little to no explanatory power in using FDI in examining producer behavior. As always, models are "as if" devices.

Free disposability of output (FDO)

FDO requires that if x can produce y, it can also produce any output bundle, y^*, whose elements, $y_m^*,\ m = 1, 2, ...M$, are no larger than the corresponding element in y. For example, suppose that 1 acre of land and 1 unit of hired labor can produce 40 bushels of corn and 40 bushels of soybeans. Then, FDO requires that 1 acre of land and 1 unit of hired labor can produce 40 bushels of corn and 0 bushels of soybeans, 0 bushels of corn and 40 bushels of soybeans, 30 bushels of corn and 30 bushels of soybeans, and so on. Producers can *dispose of (throw away) outputs* without changing input use or the production of other outputs. Figure 3 illustrates the impact of imposing FDO on the nonempty T depicted in Figure 1. If (x^0, y^0) is feasible, FDO requires that all the points on the line segment connecting (x^0, y^0) and $(x^0, 0)$ belong to T.

Production processes exist for which FDO is implausible. An important counterexample is offered by processes that involve the production of *undesirable outputs*. For example, farming that involves tilling land generates soil erosion. Soil erosion is undesirable because

Figure 3: Free disposability of output



it pollutes water resources and wastes a valuable resource, the soil. If FDO were true, farmers could hold all inputs and crop production constant and reduce soil erosion. In fact, FDO requires that soil erosion could be eliminated. But soil erosion occurs. Is it that farmers have a pathological need to pollute or to waste valuable resources? No, FDO is implausible in this setting because reducing erosion requires diverting resources from crop production. If production processes generate undesirable outputs, the rule of thumb is that resources will have to be diverted from the production of desirables to remediate the problem.

T is a convex set

Diminishing marginal returns is a fundamental concept in economics. Its essence is that one receives smaller marginal returns or benefits as one works harder or consumes more. In producer theory, its most common manifestation is *diminishing marginal productivity.* The consumer analogue is *diminishing marginal utility.*

Assuming that T is a convex set is the way that economists ensure that T is consistent with diminishing returns. A set is *convex* or satisfies *convexity* if any weighted average of any two members of the set also belongs to the set. Geometrically, a set is convex if a line segment connecting any two elements of the set lies entirely within the set (see Figure 4). The math names for the expression

$$\lambda\left(x^0, y^0\right) + (1-\lambda)\left(x^1, y^1\right)$$

for $\lambda \in (0,1)$ are the *convex combination* of (x^0, y^0) and (x^1, y^1) or the *line segment* connecting them. Thus, another way to define a set as being convex is to require that all the convex combinations of the elements of the set belong to the set. Familiar examples are the set of real numbers $(-\infty, \infty)$ and the (x, y) coordinate plane.

Let (x^0, y^0) and (x^1, y^1) in Figure 5 be two feasible input-output bundles. Convexity requires that $\lambda(x^0, y^0) + (1-\lambda)(x^1, y^1) \in T$ for all $\lambda \in (0,1)$. For example, setting $\lambda = \frac{1}{2}$ gives

$$\left(\frac{x^0 + x^1}{2}, \frac{y^0 + y^1}{2}\right) \in T, \tag{1}$$

which is the simple average of (x^0, y^0) and (x^1, y^1) that Figure 5 illustrates as falling halfway

Figure 4: Convex and Nonconvex Sets

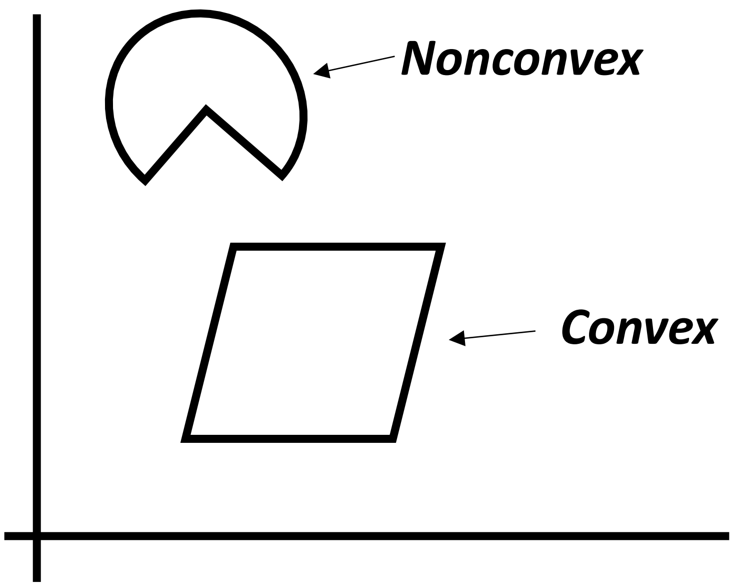
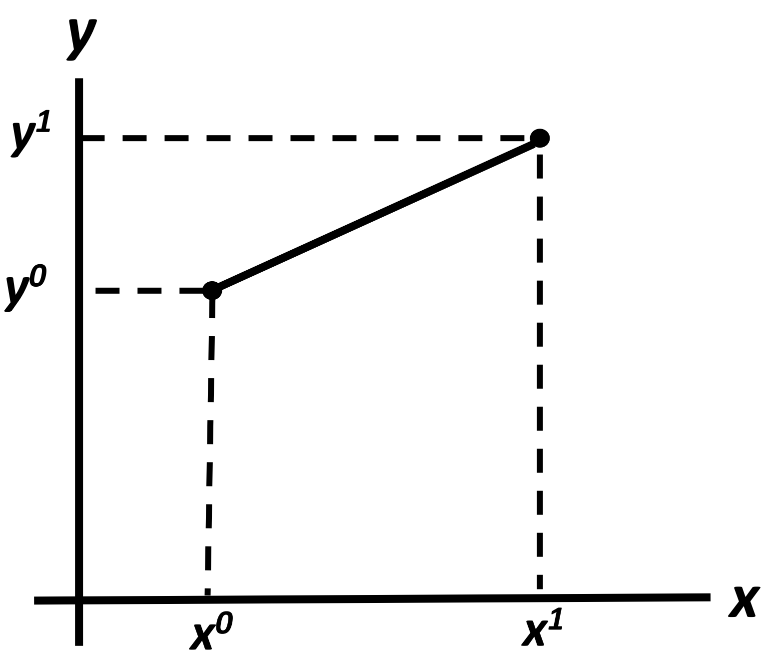


Figure 5: Convexity and T



between them on the connecting line segment. Taking $\lambda = \frac{3}{4}$ gives

$$\frac{3}{4}\left(x^0, y^0\right) + \frac{1}{4}\left(x^1, y^1\right) \in T \tag{2}$$

that corresponds to the point falling midway between $\left(\frac{x^0+x^1}{2}, \frac{y^0+y^1}{2}\right)$ and (x^0, y^0) on that line segment. Repeating this argument for all $\lambda \in (0, 1)$ shows that all the points on the line segment connecting $(x^0, y^0) \in T$ and $(x^1, y^1) \in T$ must belong to T.

Invoking convexity exposes some mental puzzles associated with working with real numbers. For example, it's sensible to think of automobile as the output of a production process. But how does one interpret

$$\frac{1+2}{2} = 1.5$$

automobiles? Is $\frac{1}{2}$ of an automobile meaningful, other than as junk? We don't provide an answer. They exist, and you are encouraged to seek them out. What's important is that you realize such questions exist.

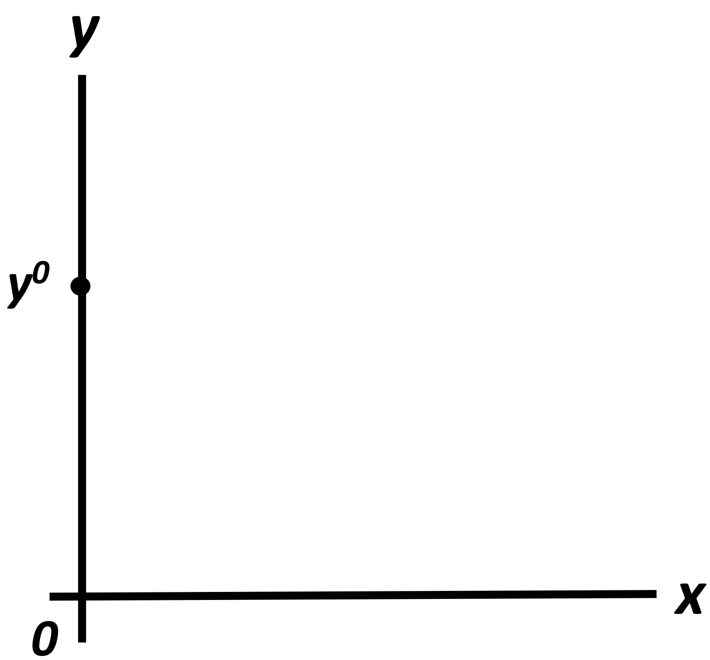
No free lunch (NFL) and No fixed costs (NFC)

NFL requires that an input bundle with all inputs equal to 0 cannot produce an output bundle that contains any positive elements, that is one with at least one $y_m > 0$. Producing an output requires using some inputs. NFL is hard to contradict. In essence, it repeats the *Law of Conservation of Mass*: matter can be neither created nor destroyed. In thinking about NFL, it helps to remember that when we say that "we make a product", we do not mean that we "create" it. Instead, we mean that we transform an existing mass of matter in the form of inputs into other forms. NFL ensures that a point such as $(0, y^0)$ in Figure 6 cannot be feasible. NFC, which is also referred to as *Inaction is possible*, permits the producer to choose to do nothing. Combining NFL and NFC ensures that points on the vertical axis above the origin cannot be feasible, while the origin remains in T.

The Canonical Technology

We say that T is a *canonical technology* if it satisfies all five assumptions. We use a distinct term for such a technology to emphasize that production processes exist that do not satisfy all

Figure 6: Violation of No Free Lunch



five assumptions. Calling it *canonical* signals that it's a standard or familiar representation.

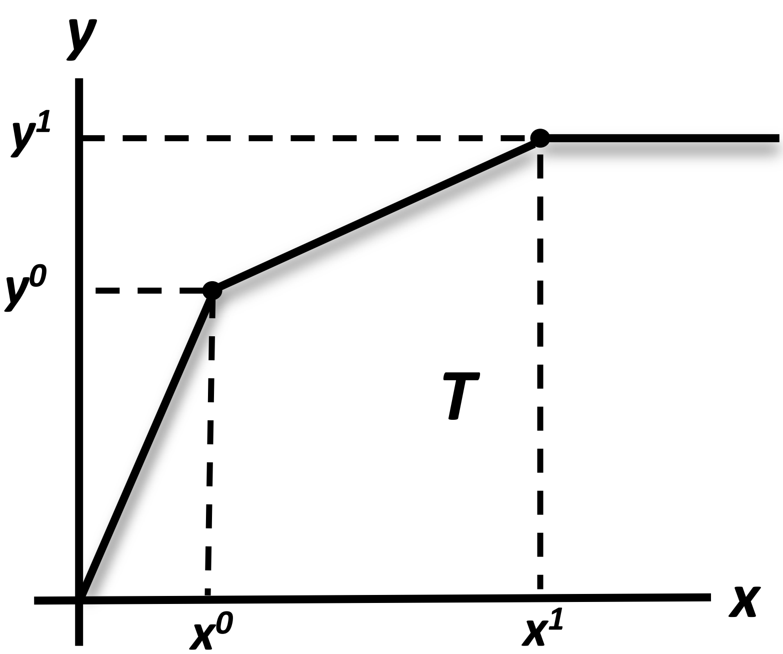
Our visual illustration of its construction highlights a key point: one can use observed data on inputs and outputs to build a set, call it $\hat{T}$, that satisfies all five assumptions. That has important implications. First, because such a $\hat{T}$ always exists, we cannot use observed data to falsify the five assumptions. That does not mean that they are correct or that we can verify them. Rather, we cannot rule them out using real-world data. Second, observed data can be used to construct a 'canonical technology'. Our *model* is not just a conceptual toy, it has empirical content and can be used to study observed phenomena.

To demonstrate the construction, we assume that we are given two data points (x^0, y^0) and (x^1, y^1) as illustrated in Figure 5. (Thus, nonemptiness is satisfied.) Invoking convexity incorporates all points on the solid line segment connecting these two points into the technology. Imposing FDO on each point on this line segment (including its endpoints) ensures that all points lying below it and on or above the horizontal axis belong to T. Applying FDI to the result includes all points lying below a horizontal line emanating from (x^1, y^1) (towards the right) and on or above the horizontal axis. Adding assumption 5 includes $(0, 0)$. Convexity includes the line segment connecting $(0, 0)$ and (x^0, y^0). FDO includes the area below that line segment. The result, as built up from (x^0, y^0) and (x^1, y^1), is illustrated by Figure 7. (In Exercise 2, you are asked to revisit this construction using an alternative approach to developing T.)

Figure 7 was constructed using only two data points and our assumptions. That accounts for its *polygonal* appearance.[4] You should not infer, however, that the canonical technology must be polygonal. Figure 7 ensures that we do not rule that out, but other forms are allowed. For example, Figure 8 illustrates T as the area below the smooth curve emanating from the origin labelled OY, which includes the points (x^0, y^0) and (x^1, y^1). You can verify that the technology illustrated in Figure 8 satisfies our five assumptions. Observing (x^0, y^0) and (x^1, y^1) is consistent with both the technology illustrated in Figure 7 and the technology in Figure 8. But the technology in Figure 8 includes points not included in the technology illustrated in Figure 7. One interpretation of the technology in Figure 7 is as a *conservative*

[4] A *polygon* is a figure in the coordinate plane with a finite number of sides. Special cases are triangles, rectangles, and squares.

Figure 7: The Canonical Technology



approximation to a technology consistent with the technology illustrated in Figure 8 (and possibly others) that has been constructed using observed (x^0, y^0) and (x^1, y^1) and our five assumptions.

Figure 8: A smooth canonical technology

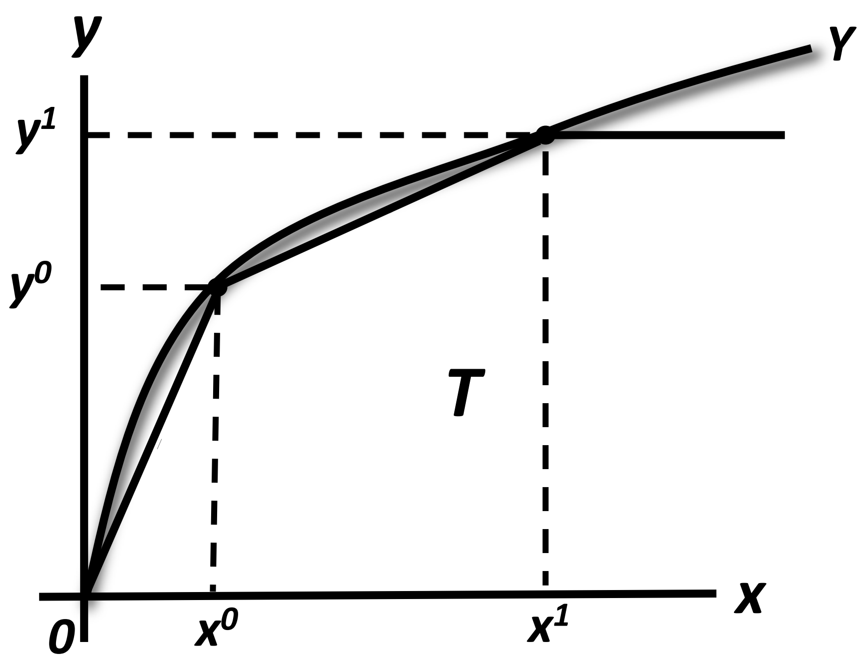
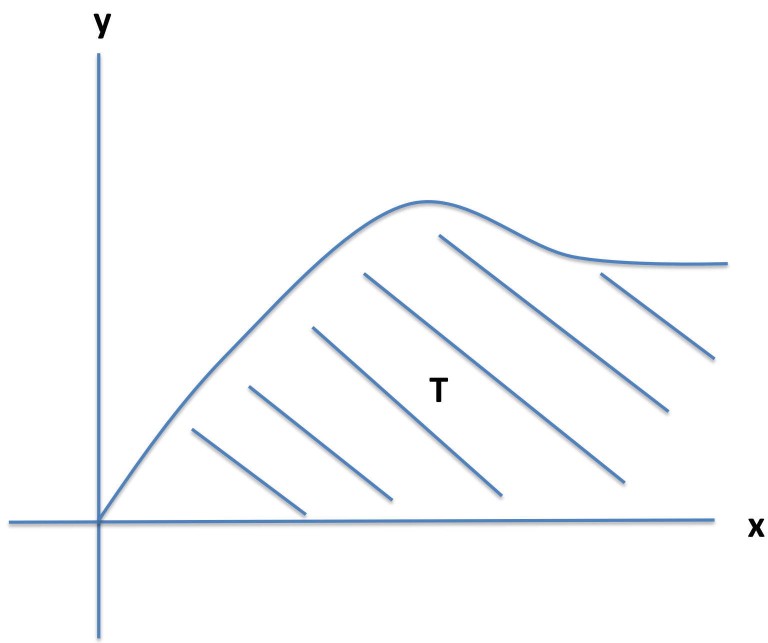


Figure 9: Exercise 1



Exercises

1. Figure 9 depicts a technology. Explain which of the properties it satisfies.

2. Assume that you are given two data points (x^0, y^0) and (x^1, y^1) as illustrated in Figure 5. Construct T by invoking in succession free disposability of inputs, convexity, free disposability of outputs, and no free lunch and no fixed costs. Compare and contrast your result with Figure 7.

Lectures on Neoclassical Production Economics[1]

Lecture 3: Production and Input Requirement Functions

Robert G. Chambers[2]

January 9, 2024

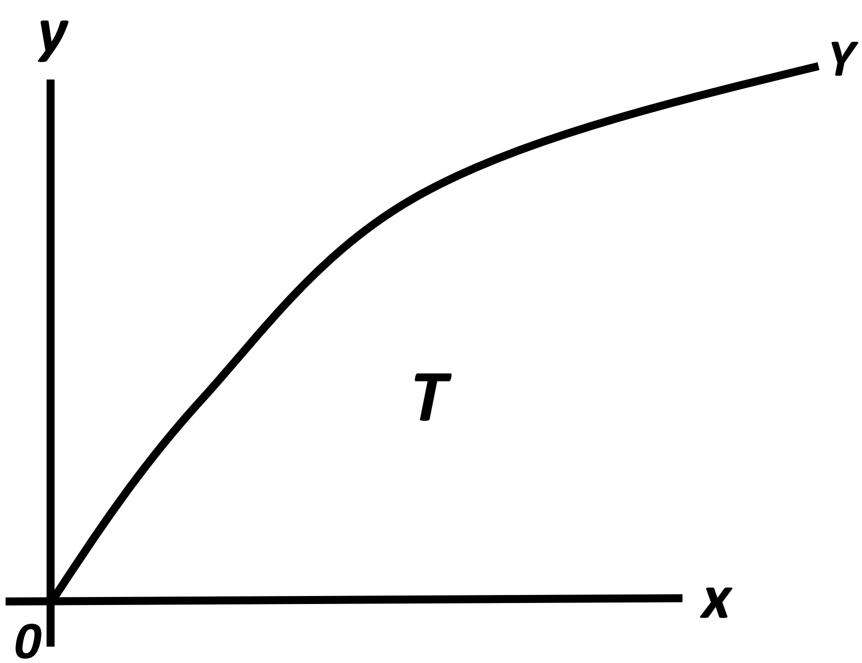
[2]University of Maryland

Lecture Introduction

So far, we have discussed the technology as a set of inputs and outputs. Many readers, however, will have encountered the technology represented as a *production function*. Some may have also encountered an *input requirement function*. This lecture shows how to derive these functions from T and how our basic assumptions on T affect their properties.

Figure 1 depicts a single-input, single-output, canonical technology, T, as everything on or below the curve OY.[1] Such a technology is unrealistic. The same is true for single-

Figure 1: Canonical Technology



input, single-output production processes. We study them because they provide the core

[1] Figure 1 is derived from Figure 8 in Lecture 2.

economic intuition on producer behavior and a nice format for showing how the notion of the technology is incorporated into economic analyses.

We first study the *production function* for a single-output, single-input production process, visualized as the curve 0Y in Figure 1. Then we study the *input requirement function*, visualized as 0X in Figure 2. We conclude the lecture by looking at the (slightly) more general cases of a multiple-input, single-output production process and a single-input, multiple-output production process.

But first, a few words on viewing phenomena from different perspectives.

A Parable on Perspective

> When three blind men were confronted with, to them, an unknown beast called an *elephant,* they decided to identify its nature by examining it by touch. The first, who touched its side, declared it to be a wall. The second, who touched its ear, declared it to be a fan. And the third, who felt its tail, proclaimed it a snake. (Source: Apochryphal)

The moral is that examining things from different perspectives conveys different information. Think, for example, what you do when handed an unfamiliar object. Do you turn it over to look at it from all sides? We can draw insight from doing the same thing, metaphorically speaking, with our model of T.

By convention, ordered pairs graphed in the real plane measure the *independent variable* along the horizontal axis (labelled x in Figure 1) and the *dependent variable* along the vertical axis (labelled y in Figure 1).[2] This captures the notion that the independent variable drives or determines the dependent variable. So, following this convention, Figure 1 suggests that choosing x fixes the range of possible outputs that can be produced. Our definition,

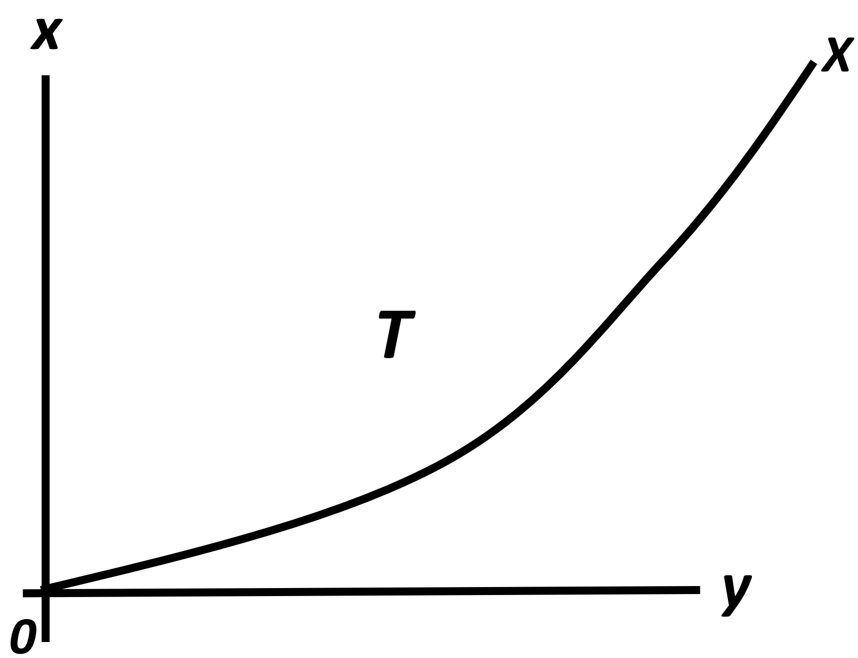
$$T = \{(x, y) \colon x \texttt{ can produce } y\},$$

written in the *active voice* with x as the subject, reinforces that intuition. But it can be useful

[2]We should say the *convention outside of economics* because economists violate this rule when they graph demand and supply curves.

to view technologies treating y as the independent variable and x as the dependent variable. This involves switching axes in depicting T. Figure 2 illustrates with T now represented as everything on or above the curve 0X. This changes the visual perspective on T, but not the

Figure 2: Canonical T in input dependent form



substance of T. We rephrase the definition of T in the *passive voice* as,

$$T = \{(x, y) : \ y \texttt{ can be produced by } x\},$$

Examining T from different perspectives is a recurring theme in these lectures. In this lecture, its manifestations are the production function and the input-requirement function. Later lectures extend these concepts to encompass more general technologies and decision settings.

Production Function (Single-input, single-output Case)

Definition 1 (The Production Function)**.** The production function for input level x, $f(x)$, is the maximum output that can be produced using x.

The mathematical version is

$$f(x) \equiv \max \{y : (x, y) \in T\}.$$

(Recall that $\equiv$ translates as *is defined.*) Here and elsewhere, we interpet $\mathtt{max}\{\cdot\}$ as an instruction to choose the maximum (largest) element contained in the set $\{\cdot\}$. In this instance, that set is all the output levels that can be produced using x, that is

$$\{y : (x, y) \in T\} = \{y : \ x \ \texttt{ can produce } \ y\}.$$

One locates $f(x)$ in Figure 1 by choosing x on the horizontal axis and then locating the largest (highest) y in T that x can produce. Doing this for all x in Figure 1 gives the curve 0Y as the *graph* of input-output pairs, $(x, f(x))$, falling on the production function for T. For the single-input, single-output technology, you can think of the graph of the production function as the *upper boundary* for T.

To appreciate an economist's interest in $f(x)$, consider the simple economic problem involving a producer who is endowed with a fixed amount of x (assumed nonconsummable). How would the producer use that x? One plausible alternative is to convert it into as much y as possible via T. $f(x)$ achieves that.

We now investigate how the basic assumptions on T affect $f(x)$.

T **is not empty** $T \neq \emptyset$ means that at least one (x, y) is feasible. Thus, *at least one x exists for which $f(x)$ is well-defined.* Here's the reasoning. If T is not empty, take any $(x, y) \in T$. If that y is the only output that x can produce, then $y = f(x)$. If other output levels can be produced using x, the largest of them is $f(x)$. Thus, $f(x)$ must exist for at least one x if $T \neq \emptyset$, as desired.

T not empty, however, *does not imply that $f(x)$ is well defined for all x.* The reasoning here is that $T \neq \emptyset$ *is satisfied even if only one feasible input-output bundle exists.* (Imagine a T that has a single point in it.) The production function is well-defined for that x, but what about other values of x?[3]

We state the conclusion in math terms as:

- $T \neq \emptyset \Rightarrow \exists\, x$ such that $f(x)$ is well defined.

In words, T not equal to the empty set implies that there exists an x for which $f(x)$ is well defined.[4]

Free disposability of inputs If FDI holds, then $(x, y) \in T \Rightarrow (x', y) \in T$ for $x' \geq x$, or, in words, increasing x does not result in a lower y. To determine its consequences, first find the production function for x, $f(x)$. By the definition of $f(x)$, x must be able to produce $f(x)$ so that $(x, f(x)) \in T$. Now substitute $f(x)$ for y in the statement of FDI to get

$$(x, f(x)) \in T \Rightarrow (x', f(x)) \in T$$

for $x' \geq x$. Because $f(x)$ can be produced using x', it cannot be true that the *maximum output* for x' can be less than $f(x)$. It must be that $f(x') \geq f(x)$. We conclude that FDI implies that the production function is *nondecreasing* in the input,

- $FDI \Rightarrow f(x') \geq f(x)$ for $x' \geq x$ (nondecreasing in x).

The intuitive idea is that increasing input use causes maximum output, $f(x)$, to increase. So, the graph of $f(x)$ has positive slope as x changes. 0Y in Figure 1 illustrates. While correct, that intuition is not precise. Under FDI, an increasing $f(x)$ is possible but not required. What FDI requires is that $f(x)$ not fall as input rises, or that the graph of $f(x)$ not have a negative slope.

[3] There are mathematical conventions ('tricks') that we can use to circumvent this problem. For example, one such trickis to set $f(x)$ to a tiny nonnegative number. That solves the mathematical problem, but that makes for problematic intuition. For example, if we set $f(x) = 0$ in such instances, that implies that the physical input disappears, leaving the unhappy suggestion that T violates the Law of the Conservation of Mass.

[4] Math Note: Here we implicitly invoke the regularity condition imposed on T to ensure the existence of a well-defined maximum.

Remark 1. *Understanding the difference between the intuitive meaning of a statement and its precise meaning is important. But my belief is that your primary focus should be on the intuition. If you have done that, you can refine that intuition to be more precise. The terminology "increasing" and "nondecreasing" in an input offers a good example. The first means that as the input increases output goes up. The second means that it does not go down. Going up is a special case of not going down, as is staying constant. But the "intuition" here is that output increases.*

Free disposability of output Under FDO,

$$(x, y) \in T \Rightarrow (x, y') \in T$$

for $y \geq y'$. For any $(x, y) \in T$, the definition of $f(x)$ ensures that $f(x) \geq y$. On the other hand, $(x, f(x))$ must belong to T. Applying FDO requires that x can also produce any y less than or equal to $f(x)$, so that $f(x) \geq y$ implies that $(x, y) \in T$.

Thus, if FDO applies, knowing $f(x)$ lets one construct T. But if one knows T, one can construct $f(x)$ for single-output technologies. We, therefore, conclude that under FDO $f(x)$ and T convey the same information. That is,

- $FDO \Rightarrow\ (x, y) \in T \Leftrightarrow f(x) \geq y,$

which translates as: the statements that $(x, y) \in T$ and that $f(x) \geq y$ are equivalent. The production function, $f(x)$, is a *function representation of T*.[5]

T is a convex set Convexity requires that weighted averages of two feasible points must be feasible. In mathematical form,

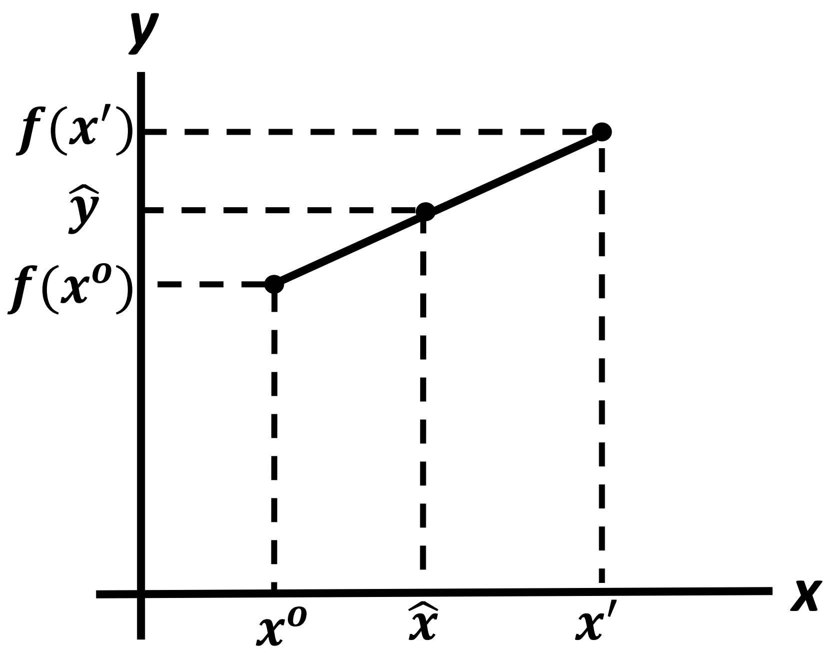
$$(x^o, y^o) \in T \texttt{ and } (x', y') \in T \Rightarrow \lambda (x^o, y^o) + (1 - \lambda)(x', y') \in T$$

for $\lambda \in (0, 1)$.

[5]Many, if not most, elementary texts define the "technology" by the production function without invoking or recognizing FDO. For simple enough cases, that gives a workable model. But the approach becomes more problematic as more general production processes are considered.

To investigate the consequences for $f(x)$, choose $y^o = f(x^o)$ and $y' = f(x')$. Both $(x^o, f(x^o))$ and $(x', f(x'))$ fall on the graph of $f(x)$, the curve 0Y in Figure 1. Imposing onvexity ensures that all weighted averages of $(x^o, f(x^o))$ and $(x', f(x'))$ belong to T. These weighted averages are illustrated in Figure 3 as the points on the line segment connecting $(x^o, f(x^o))$ and $(x', f(x'))$. Pick any $\lambda \in (0,1)$ and define

Figure 3: Convexity and the production function



$$\hat{x} = \lambda x^o + (1 - \lambda) x',$$

which we illustrate in Figure 3 as falling between x^o and x'. We seek

$$f(\hat{x}) = f(\lambda x^o + (1 - \lambda) x').$$

Convexity implies that $(\hat{x}, \hat{y}) \in T$, where

$$\hat{y} = \lambda f(x^o) + (1-\lambda) f(x').$$

(See Figure 3.) It must be true, therefore, that the largest output that $\hat{x}$ can produce, $f(\hat{x})$, cannot be smaller than $\hat{y}$. Therefore,

$$\begin{aligned} f(\lambda x^o + (1-\lambda) x') &= f(\hat{x}) \\ &\geq \hat{y} \\ &= \lambda f(x^o) + (1-\lambda) f(x') \end{aligned}$$

In Figure 3, either $\hat{y} = f(\hat{x})$ or $(\hat{x}, f(\hat{x}))$ lies above $(\hat{x}, \hat{y})$. If it's the latter, the graph of $f(x)$ passes above the line segment connecting $(x^o, f(x^o))$ and $(x', f(x'))$. Because our choice of λ was arbitrary, the same argument applies for all $\lambda \in (0, 1)$. That yields the conclusion:

- `Convexity` $\Rightarrow f(\lambda x^o + (1-\lambda) x') \geq \lambda f(x^o) + (1-\lambda) f(x')$ for $\lambda \in (0,1)$.

$f(x)$ is a *concave function* of x.[6] The curve 0Y in Figure 1 depicts the graph of an increasing and concave function of x. The positive slope comes from FDI, and the behavior of the slope as x varies is a consequence of convexity. In Figure 1, the slope of 0Y gets flatter as x increases. This is a characteristic of concave functions.[7] Thus, together FDI and convexity (speaking loosely) require that the *production function increases at a decreasing rate as x increases* and manifests the *law of diminishing returns*. We defer a more complete discussion of diminishing returns until after we discuss the consequences of NFL and NFC for $f(x)$.

No Free Lunch and NFC NFL requires that $(0, y) \notin T$ if $y \geq 0$ and $y \neq 0$. For the single output case, we rewrite that as $(0, y) \notin T$ if $y > 0$. A positive output cannot be produced using zero input. $f(0)$ cannot be greater than zero. Adding NFC requires that $(0, 0) \in T$ leading to the conclusion that $f(0) = 0$. Thus,

- NFL `and` $NFC \Rightarrow \ f(0) = 0$

NFL and NFC imply that the graph of the production function emanates from the origin, again as illustrated in Figure 1 by the curve $0Y$.

[6] A function $h(z)$ is defined to be concave in z if $h(\lambda z^o + (1-\lambda) z') \geq \lambda h(z^o) + (1-\lambda) h(z')$ for $\lambda \in (0,1)$.

[7] More precisely, a concave function's slope does not increase as x increases.

Average Products, Marginal Products, and the Law of Diminishing Returns

We use the production function to define and investigate the concepts of *average product* and *marginal product.*

Definition 2 (Average Product)**.** The average product at input level $\hat{x}$ is the ratio of $f(\hat{x})$ to $\hat{x}$

In equation form,

$$AP(\hat{x}) \equiv \frac{f(\hat{x})}{\hat{x}},$$

where AP stands for average product. Because $f(x)$ changes as x varies, the average product is a function of x. In Figure 4, $AP(x^0)$ is the vertical distance $0f(x^0)$ divided by the horizontal distance $0x^0$, which equals the slope of the dotted line segment connecting the origin, 0, and $(x^0, f(x^0))$.

In Figure 4, $AP(x^0) > AP(x^1)$. The average product falls as x increases. You can experiment with modifications of Figure 4 to convince yourself that this is true for the canonical technology. Economists call this phenomenon a *diminishing average product* or *diminishing average returns.* It is a consequence of imposing convexity on T, and it implies that $f(x)$ grows, in percentage terms, at a slower rate than x

Definition 3 (Arc Marginal Product)**.** The arc marginal product is the ratio of the change in the value of $f(x)$ to the change in x as x varies.

A standard way to write this definition is

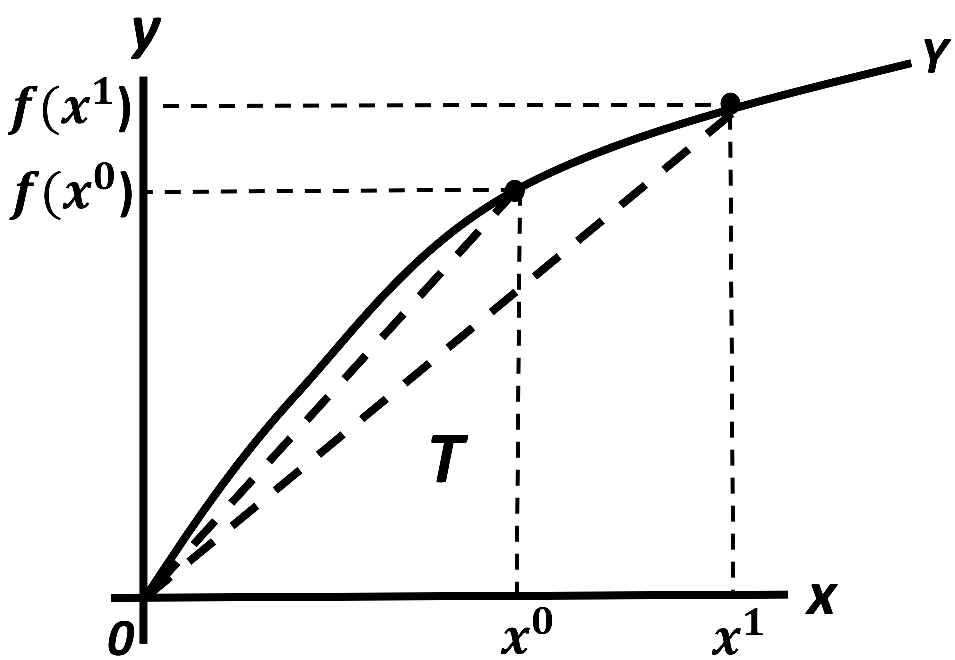
$$MP(x) \equiv \frac{\Delta f(x)}{\Delta x}$$

where MP stands for marginal product and capital Greek delta (Δ) denotes change. Starting at $(x^0, f(x^0))$ and moving to $(x^1, f(x^1))$, we calculate that ratio in Figure 4 as

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x^1) - f(x^0)}{x^1 - x^0},$$

which is given by the slope of the line segment connecting $(x^0, f(x^0))$ and $(x^1, f(x^1))$ (not drawn) in Figure 4.

Figure 4: Average Product



On the meaning of 'marginal'

The adjective *marginal* recurs repeatedly in economic discussions. Its use in the English literature was popularized by Alfred Marshall, who attributed it to a German term used by Johann Heinrich von Thünen. he *Oxford English Dictionary* offers the following definition (among others):

> Of or relating to the extra cost, revenue, or utility involved in or deriving from the production or consumption of one additional unit of a product.

We paraphrase this to suit our discussion as

> Of or relating to the extra production deriving from the use of one additional unit of a input.

Note that the statement "one additional unit" leaves unpecified the units in which x and y are measured. Think of our agricultural illustrations and let y stand for corn and x for land. Is corn measured in bushels or metric tonnes? Do we measure land in acres, hectares, square miles, or square kilometers? Units matter. Change the units and the measures change. The overwhelming tendency in economics is to interpret 'marginal' changes as involving 'tiny' units so that $x^1 - x^0$ is practically zero (infinitesimally small).

Speaking in 'tiny-units' terms, the marginal product *often* has a simple *geometric intepretation*.

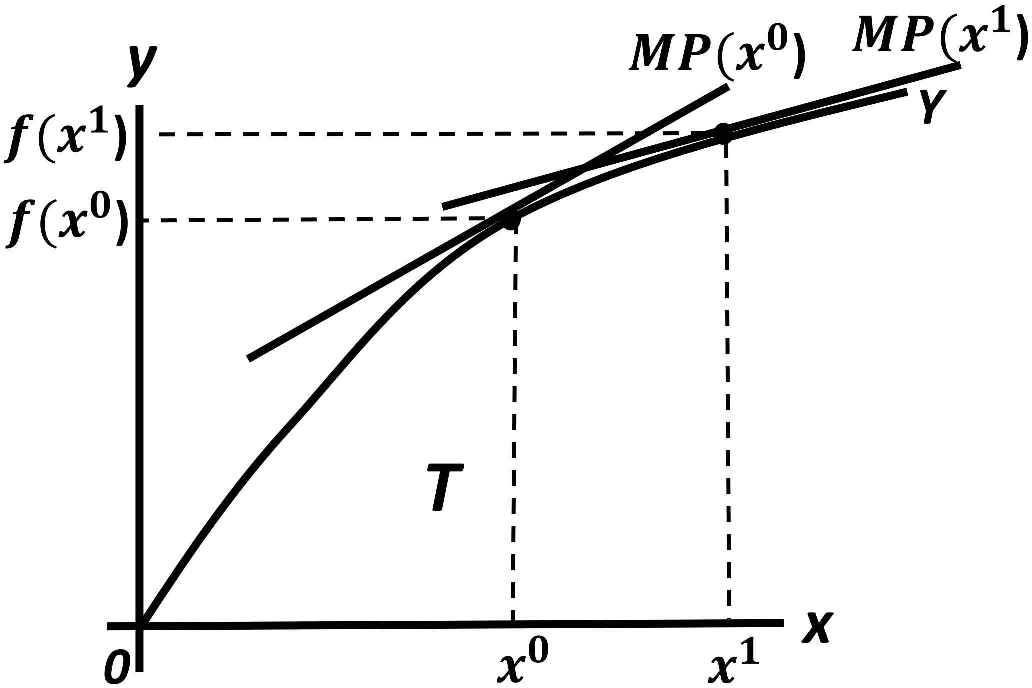
Definition 4 (Marginal Product)**.** The marginal product at x^0 is the slope of the line segment tangent to the graph of $f(x)$ at $(x^0, f(x^0))$. [8]

Figure 5 illustrates for x^0 and x^1. The respective marginal products are given by the slope of the tangent line segments $MP(x^0)$ and $MP(x^1)$. Observe in Figure 5 that $MP(x^0) > MP(x^1)$, the slope of $f(x)$ decreases as x increases. Thus, $f(x)$ exhibits a *diminishing marginal product* in x.

A diminishing marginal product means that x becomes less effective at promoting extra output growth as its use grows. Hence, successively applying the same dose of x brings

[8] For you calculus cowboys, we take $lim\left(x^1 - x^0\right) \longrightarrow 0$.

Figure 5: Marginal Product



forth less output with each dose, a phenomenon that economists have observed as *eventually occurring* for many production processes.[9]

The production functions depicted in Figures 4 and 5 exhibiting both a *diminishing average product* and a *diminishing marginal product* is a consequence of imposing convexity on T. Figure 6 illustrates another consequence of convexity. There $AP\left(x^0\right)$ and $MP\left(x^0\right)$ are depicted together, and it is apparent that

$$AP\left(x^0\right) \geq MP\left(x^0\right).$$

Experimenting with Figure 6 confirms that this inequality holds for all x for a canonical technology.

Defining the marginal product as the slope of the line segment tangent to the graph of the production function is problematic if its graph is "kinked". Figure 7 in Lecture 2 illustrates. At either (x^0, y^0) or (x^1, y^1) the graph of $f(x)$ for the illustrated T has multiple (literally an infinity) tangent line segments. The multiple tangencies manifest the difference, for example, between the arc marginal products calculated at (x^1, y^1) for increases and decreases in x. For increases, the arc marginal product is zero, but for decreases it's the slope of the line segment connecting (x^0, y^0) and (x^1, y^1).[10]

You may have encountered production functions that look different than $f(x)$ in Figure 1. Many discussions depict the production function as resembling the curve labelled 0Y in Figure 7, which violates convexity. Under FDO, T includes all points falling on or below 0Y. Connect a point such as A and the origin with a straight line segment (not drawn). If T were convex, points on the line segment should fall in T. But they don't. Hence, this technology is not convex. You will also note that for the technology in Figure 7 both $AP(x)$ and $MP(x)$ first increase and then decrease. (You are asked to verify this visually in Exercise 1.) Production processes like the ones in Figure 7 are said to exhibit *variable returns* to distinguish them from convex ones. Such production processes can and do exist, but they are not a focus of these lectures. We explain why in Lecture 9.

[9] von Thünen was among the first to document it empirically, and he did it in the early 19th century using German farming data.

[10] In analytic terms, at $\left(x^1, y^1\right)$ the left-hand derivative of $f(x)$ differ, so that the function is not differentiable.

Figure 6: Average Product and Marginal Product

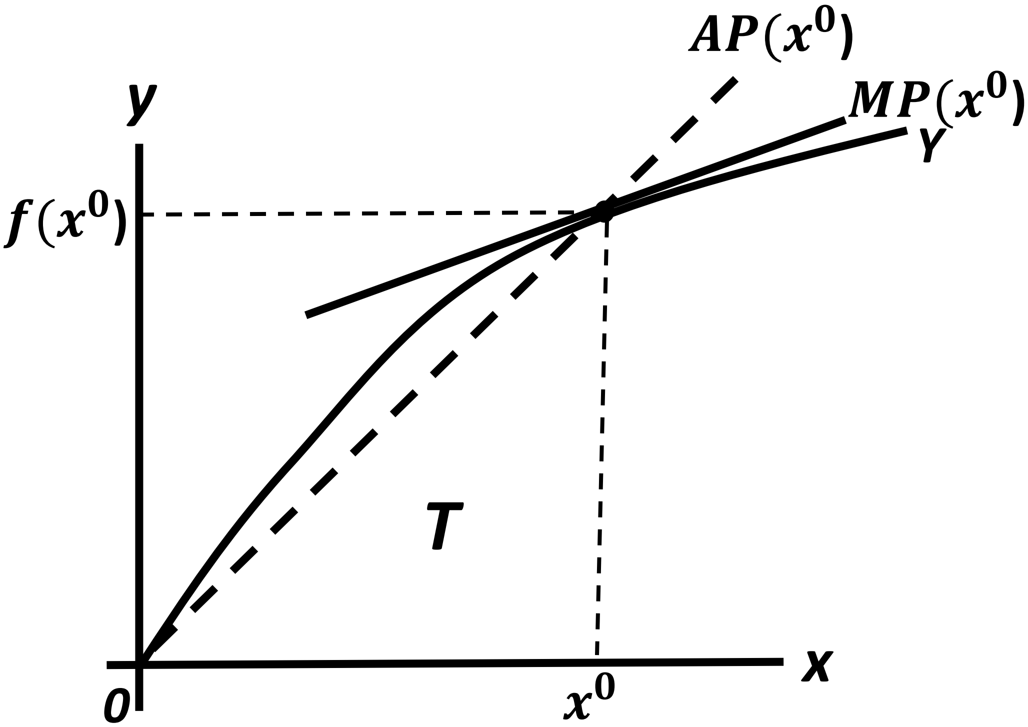
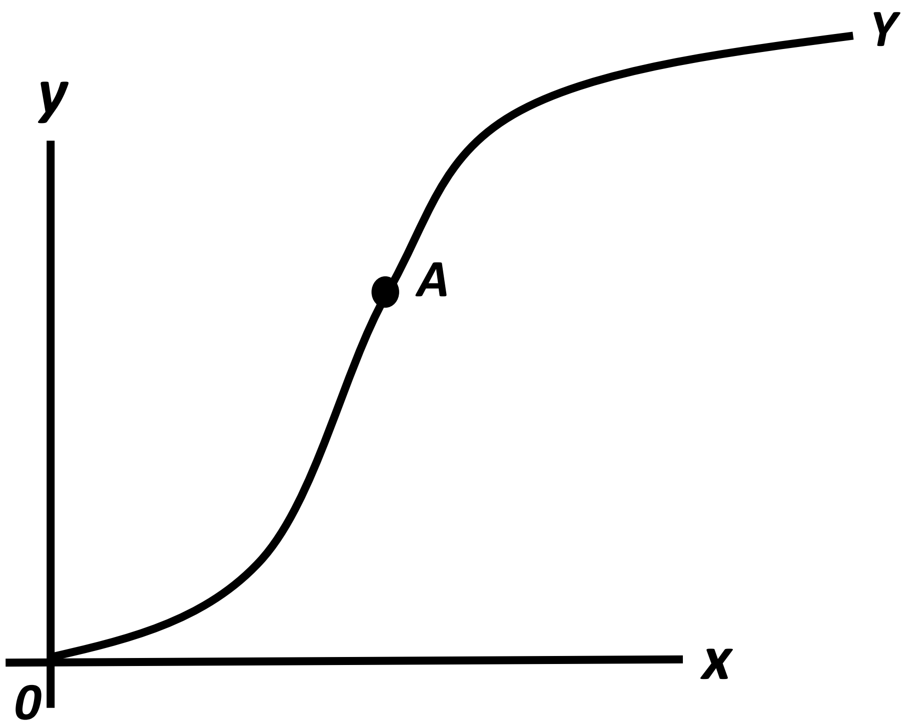


Figure 7: Variable Returns Production



Input-requirement Function (Single-Input, Single-Output Case)

Definition 5. [Input-requirement Function] The input-requirement function for output level y is the smallest input that can produce y.

The mathematical definition is:

$$e(y) \equiv \min \{x : \ x \ \texttt{can produce} \ y\}.$$

Here and elsewhere, the expression $\texttt{min}\{\cdot\}$ is an instruction to choose the smallest (minimal) element in the set $\{\cdot\}$. If T satisifies FDO, Definition **??** is equivalent to

$$e(y) \equiv \min \{x : \ f(x) \leq y\}.$$

Here we use the fact FDO ensures that $f(x)$ is a function representation of T.

We locate $e(y)$ in Figure 2 by choosing y and then locating the smallest (lowest) x in T that can produce y. This gives the curve 0X as the graph of the input-output pairs $(e(y), y)$ that lie on the input-requirement function for T.[11]

A thought experiment reveals why an economist might be interested in $e(y)$. Consider a self-interested producer, facing a single-input, single-output technology, who is contractually obligated to deliver a fixed amount of y. How much x would that producer use in producing that output?

We now develop the consequences of our basic assumptions for $e(y)$.

T is not empty We have:

- $T \neq \emptyset \Rightarrow \exists \, y : e(y)$ `is well defined.`

In words, T not equal to the empty set implies that there exists a y such that its input-requirement function, $e(y)$, is well-defined. We demonstrate using arguments that parallel

[11]Because Figure 2 depicts output, y, on the horizontal axis and x on the vertical axis, it is more *conventional* to refer to the points on the graph of the input-requirement function as output-input pairs and write them as $(y, e(y))$ in place of $(e(y), y)$. In an attempt to avoid confusion, we adhere here to the convention maintained elsewhere in these lectures of listing inputs first and outputs second.

our earlier demonstration that $T \neq \emptyset$ implies the existence of at least one well-defined production function. Recall that if T is not empty, at least one feasible input-output pair, (x, y), must exist. If that x is the smallest that can produce y, then $x = e(y)$. If it's not, $e(y)$ will be smallest that can produce that y. That shows that one y exists for which $e(y)$ is well-defined. But just as $T \neq \emptyset$ did not imply the existence of a production function for all x, $T \neq \emptyset$ does not imply the existence of an input-requirement function for all y. And, again, the reason is that $T \neq \emptyset$ is satisfied if only one feasible input-output pair exists.

Free disposability of input The consequence of imposing FDI for $e(y)$ is that

- $FDI \Rightarrow\ x \geq e(y) \Leftrightarrow (x, y) \in T$,

which reads as FDI implies that $x \geq e(y)$ and $(x, y) \in T$ are equivalent statements. Thus, FDI ensures that $e(y)$ is a *function representation of* T so that knowing it is equivalent to knowing T. In Figure 2, this consequence validates the interpretation of T as all points lying on or above 0X. To verify, choose $(x, y) \in T$. Then, by the definition of $e(y)$, $x \geq e(y)$. On the other hand, because $(e(y), y) \in T$, FDI requires $x \geq e(y) \Rightarrow (x, y) \in T$.

Free disposability of output FDO implies that $e(y)$ is a nondecreasing function of y, or

- $FDO \Rightarrow e(y) \geq e(y')\ \ \texttt{for}\ \ y \geq y'$

To verify this claim, find $e(y)$ for y. Then because FDO requires that

$$(x, y) \in T \Rightarrow (x, y') \in T$$

for $y \geq y'$ substituting $e(y)$ for x gives

$$(e(y), y) \in T \Rightarrow (e(y), y') \in T$$

for $y \geq y'$. Because $e(y)$ can produce y', the smallest x that can produce y' cannot be larger than $e(y)$ which establishes the desired property. Figure 2 captures this consequence by depicting the graph of $e(y)$ (the curve 0X) with a positive slope. Strictly speaking, that graphical depiction means that $e(y)$ is increasing in y, which is permitted but not required by FDO. What FDO rules out is the graph of $e(y)$ having a negative slope.

Convexity The slope of OX in Figure 2 increases as y increases. The positive slope is a consequence of FDO. The increasing positive slope is a characteristic of a *convex function*[12] and is a consequence of imposing convexity on T. In mathematical terms,

- `Convexity` $\Rightarrow$ $e(\lambda y^o + (1-\lambda)y') \leq \lambda e(y^o) + (1-\lambda)e(y')$

for $\lambda \in (0,1)$. Convexity of T thus implies that $e(y)$ is a *convex function of y*. The demonstration parallels our earlier demonstration that imposing convexity on T implies that $f(x)$ is concave in x. (As such, we leave it to you to carry out the visual illustration of the arguments in Exercise 2.)

Pick y^o and y' and find $e(y^o)$ and $e(y')$. We seek $e(\lambda y^o + (1-\lambda)y')$ for $\lambda \in (0,1)$. If T is convex,

$$(e(y^o), y^o) \in T \texttt{ and } (e(y'), y') \in T \Rightarrow (\lambda e(y^o) + (1-\lambda)e(y'), \lambda y^o + (1-\lambda)y') \in T$$

for $\lambda \in (0,1)$. Thus, because $\lambda y^o + (1-\lambda)y'$ can be produced using $\lambda e(y^o) + (1-\lambda)e(y')$, the smallest input that can produce $\lambda y^o + (1-\lambda)y'$ cannot be greater than $\lambda e(y^o) + (1-\lambda)e(y')$, so that

$$e(y^o) + (1-\lambda)e(y') \geq e(\lambda y^o + (1-\lambda)y'),$$

as required.

A cautionary remark is merited. We use the same adjective *convex* to describe two different mathematical objects. One is a *convex set*. The other is a *convex function*. The concepts are related, and both involve taking weighted averages. But they still describe different mathematical objects that correspond to different, albeit related, economic concepts.

The economic implication of convexity is that increasing output by the same amount requires increasingly larger doses of x. This is, so to speak, the 'flip side' of a diminishing marginal product and is a precursor to the notion of *increasing marginal cost* developed in Lecture 6.

No Free Lunch and No Fixed Cost For the single output case, NFL requires that $(0,y) \notin T$ if $y > 0$. That requires that $e(y) \neq 0$ for $y > 0$. The minimal input required to

[12] A function $h(z)$ is defined to be convex in z if $h(\lambda z^o + (1-\lambda)z') \leq \lambda h(z^o) + (1-\lambda)h(z')$ for $\lambda \in (0,1)$. Another way of saying this, which you can verify yourself, is that $h(z)$ is convex if $-f(z)$ is concave.

produce a positive output cannot be zero. NFC, on the other hand, requires that $(0, 0) \in T$ allowing us to conclude that $e(0) = 0$. This is illustrated by the graph of the input-requirement function, 0X, emanating from the origin in Figure 2 and never intersecting the horizontal axis for $y > 0$.

Multiple-input, Multiple-output Technologies

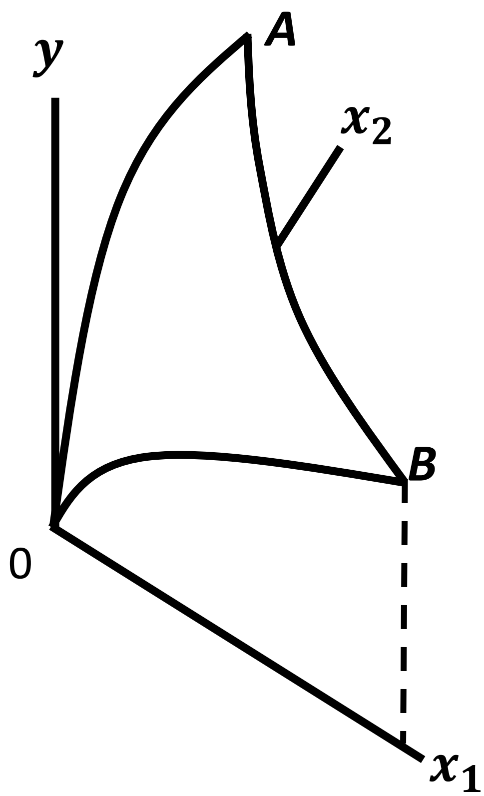
The production function is a well-defined concept for single-output, multiple-input technologies. On the other hand, in a multiple-input setting, the input-requirement function becomes definitionally suspect. When there are multiple-inputs, $x_1, x_2, ..., x_N$, one can always define *an* input-requirement function for one of the N inputs as the minimum amount of that input required to produce a fixed output bundle given that the other N-1 inputs are held constant. But the selection of the input to be minimized is arbitrary, so that it follows that N distinct input-requirement functions can be defined. Similarly, while an input-requirement function is well-defined in the single-input, the multiple-output setting, the production function concept becomes suspect because M distinct production functions can be defined.

Treating either the single-output, multiple-input production function or the single-input, multiple-output production function requires adjusting terminology and our graphical illustrations. Figure 8 depicts a single-output, two-input technology using the manifold 0AB0 emanating from the origin. In that setting, we interpret curves such as 0Y in Figure 1 as depicting how $f(x)$ varies with a single input while holding the other input constant. In the three-dimensional case, that entails, for example, projecting the graph of $f(x_1, x_2)$ from (x_1, x_2, y) space parallel to the x_2 axis onto the (x_1, y) plane. If we fix x_2 at, say $\hat{x_2}$, then Figure 1 (and others) are then interpreted as the graph on the (x_1, y) plane of $(x_1, f(x_1, \hat{x_2}))$. (See Figure 8.)

Related questions are. Which average product? Which marginal product? Is the average (marginal) product for x_1 or x_2? The relevant notions now refer to a *productivity of an input* that allows one input to vary while holding all others constant. For example, the *average productivity of* x_1, with x_2 held constant at $\hat{x_2}$ is

$$\frac{f(x_1, \hat{x_2})}{x_1}.$$

Figure 8: Production Function with Two Inputs



The *marginal productivity of* x_1 with x_2 held constant at $\hat{x_2}$ is

$$\frac{\Delta f\left(x_1, \hat{x_2}\right)}{\Delta x_1}.$$

The graphical descriptions of these productivities are as before except that the graph of $\left(x_1, f\left(x_1, \hat{x_2}\right)\right)$ or its appropriate generalization when there are more than two inputs, which is now called *the productivity response contour for* x_1 *with* $x_2 = \hat{x_2}$, or a *productivity contour*, replaces the graph of the single-output, single-input production function. Parallel adjustments are needed to accommodate the single-input, multiple-output, input-requirement function.

Exercises

1. Demonstrate visually that the technology depicted in Figure 7 exhibits initially increasing and then decreasing average products and marginal products. What characterizes the input level where AP(x) reaches its maximum.

2. Develop a graphical illustration of the consequences of convexity of T for an input-requirement function.

3.* By Definition 1, the graph of the production function, $f(x)$, in (x, y) for the technology illustrated in Figure 1 is the curve labelled 0Y. That T is convex and the resulting production function is *concave*. That is,

$$f(\lambda x^o + (1-\lambda)x') \geq \lambda f(x^o) + (1-\lambda) f(x')$$

for $\lambda \in (0,1)$. This algebraic notion of concavity can be linked more intuitively to the geometric notion of a convex set using the *hypograph* of the function, $f(x)$. The *hypograph* of the function consists of the (x, y) falling on or below the graph of $f(x)$, that is,

$$\texttt{hypograph of f} = \{(x,y) : f(x) \geq y\}.$$

In Figure 1, the hypograph consists of the points on or below the curve 0Y.

The notion of the hypograph can be used to frame a geometric definition of a concave function that connects it to set convexity. Namely, $f(x)$ *is concave if and only if its hypograph is a convex set.* Show that if

$$f(\lambda x^o + (1-\lambda)x') \geq \lambda f(x^o) + (1-\lambda) f(x')$$

for $\lambda \in (0,1)$, then hypograph of f is a convex set.

4.* A function, $e(y)$ is convex if and only if its *epigraph* is a convex set, where

$$\texttt{epigraph of e} = \{(y,x) : e(y) \leq x\},$$

consists of all the points on or above the graph of $e(y)$. Show that if epigraph of $e(y)$ is a convex set then,

$$e(\lambda y^o + (1-\lambda)y') \leq \lambda e(y^o) + (1-\lambda) e(y')$$

for $\lambda \in (0,1)$.

Lectures on Neoclassical Production Economics[1]

Lecture 4: Input Sets and Output Sets

Robert G Chambers [2]

January 9, 2024

[2]University of Maryland

Lecture Introduction

Lecture 3 treats a single-input, single-output technology that provides an unrealistic, but simple, setting for examining production relationships. The emphasis, respectively, is on how input availability drives output produced and how output needs drive input use. That analysis yields the concepts of a production function, an input-requirement function, an average product, and a marginal product. Each is basic to the way economists think about production relations.

Because it is unrealistic, it is natural to generalize. The last section of Lecture 3 briefly discussed two generalizations: a single-output, multiple-input technology and a multiple-output, single-input technology. Figure 1 depicts the graph in (x_1, x_2, y) space of a production function for a canonical single-output, two-input technology as the manifold 0AB0. Figure 2 depicts the graph of the input-requirement function for a canonical single-input, two-output input-requirement function by the manifold 0CD0. It is understood that all points falling below 0AB0 and above 0CD0 belong to the respective T.

We make no claims for realism for either of these technologies. Indeed, both are best viewed as simple models that illustrate different aspects of production technologies. For example, in Figure 1 the surface 0B depicts the graph in (x_1, x_2, y) space of the productivity contour for x_1 (holding $x_2 = 0$). The surface 0A plays a similar role in (x_1, x_2, y) space for the productivity contour for x_2 (holding $x_1 = 0$). More generally, any point on the manifold 0AB0, say $(\hat{x_1}, \hat{x_2}, \hat{y})$, will fall on two productivity contours: one for x_1 (holding $x_2 = \hat{x_2}$) and one for x_2 (holding $x_1 = \hat{x_1}$). Marginal productivities and average productivities are available for all such points using the methods in Lecture 3. Figure 2 lets us examine the interplay between the single input and the two outputs.

What Figure 1 and Figure 2 also capture, but that we have not yet discussed, is the interplay between inputs (the surface AB in Figure 1) and the interplay between outputs (the surface CD in Figure 2). This lecture studies these interactions. As before, we typically illustrate our ideas using the real plane. But, unlike Lecture 3, the analysis is for technologies with multiple inputs and multiple outputs.

Figure 1: Two-input production function

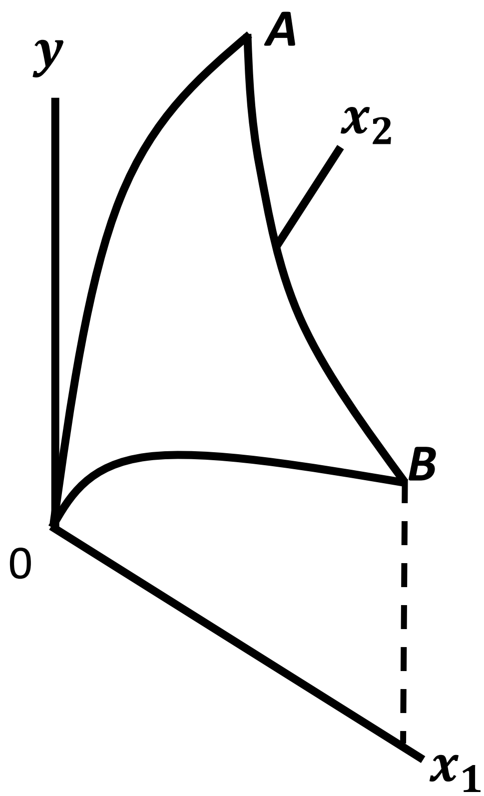
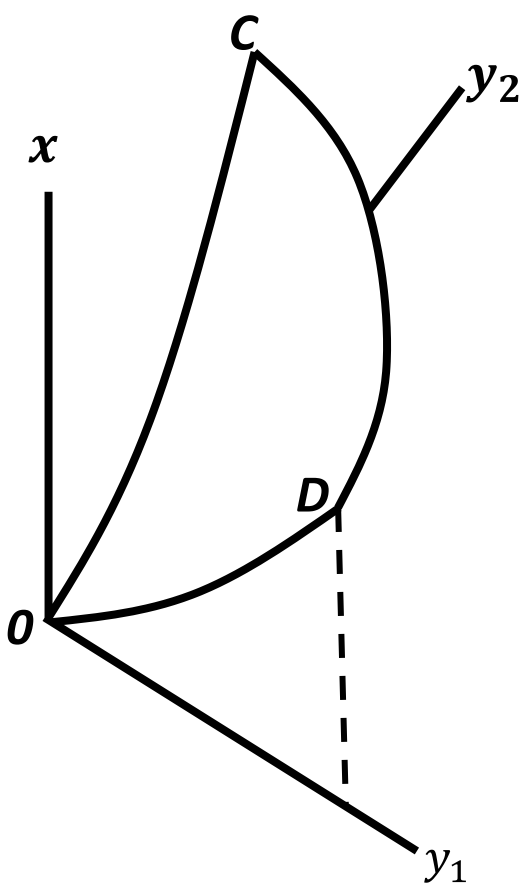


Figure 2: Two-output input-requirement function



Input Sets

Definition 1. The input set, $X(y)$, for y is the set of inputs that can produce the output vector y.

We write $X(y)$'s definition mathematically as

$$X(y) \equiv \{x : x \text{ can produce } y\},$$

or

$$X(y) \equiv \{x : (x, y) \in T\}.$$

We emphasize two aspects of $X(y)$. First, it is defined for a given output bundle, y. Therefore, y plays the role of an *independent variable*. Once it is fixed, it determines the *dependent variables*, the inputs. Apart from the presence of multiple inputs, it looks similar to the input-requirement function.[1] If y changes, $X(y)$ is expected to change. For the single-input, multiple-output case under FDI

$$X(y) = \{x : x \geq e(y)\},$$

so that $e(y)$ forms the lower boundary for $X(y)$. Second, the following equivalence relations hold

$$x \in X(y) \Leftrightarrow (x, y) \in T \Leftrightarrow x \text{ can produce } y.$$

These equivalence relations are key to determining the implications of the five basic assumptions for input sets. Our discussion provides a new visual perspective on what the assumptions mean. But the basic argument just invokes the equivalence relation $x \in X(y) \Leftrightarrow (x, y) \in T$ to recast concepts defined in terms of T in terms of $X(y)$.

T not empty $T \neq \emptyset$ means that at least one feasible (x, y) bundle exists. Because that x can produce that y, $X(y)$ for that y is not empty. That establishes that one nonempty $X(y)$ exists. Hence,

- $T \neq \emptyset \Rightarrow \exists\, y : X(y) \neq \emptyset$.

[1] For those interested in more formality, $X(y)$ is what mathematicians refer to as a *correspondence*, or sometimes as a set-valued function.

In words, T not equal to the empty set means a y exists for which $X(y)$ is not the empty set. It does not mean, however, that $X(y) \neq \emptyset$ for all possible y.

On the meaning of $\geq$ and $\leq$

This lecture focuses on comparisons of points in the coordinate plane. However, instead of following the usual convention of labelling the horizontal axis x and the vertical axis y, we will be labeling the horizontal axis as x_1 and the vertical axis x_2 in one instance and as y_1 and y_2, respectively, in another. That's because the focus is upon comparing, respectively, input and output bundles.

In making such comparisons, we need to be careful about how we interpret the math symbols $\geq$ and $\leq$. When we compare two real numbers, a and b, the meaning is clear. Then $a \geq b$ means that a is not smaller than b. Similarly, $a \leq b$ means that a is not larger b. Our ability to make such comparisons changes, however, when we move to bundles (vectors) of inputs or outputs. To take an example, suppose that $x^A = (4\ \texttt{rakes}, 2\ \texttt{shovels})$ and $x^B = (2\ \texttt{rakes}, 4\ \texttt{shovels})$. Can one meaningfully say that either $x^A \geq x^B$ or that $x^B \geq x^A$? The answer is no, because A has more rakes than B but less shovels.

So, all points in the coordinate plane (or its higher-dimensional generalizations) cannot be ranked (ordered) relative to one another using either $\geq$ or $\leq$. $\geq$ and $\leq$ are examples of *partial orderings* of the coordinate plane. It's not important that you remember this "math" term because we won't use it again. But is important to remember that our ability to compare points in the coordinate plane using $\geq$ and $\leq$ is limited. That, in turn, means that properties of T that hinge upon $\geq$ or $\leq$ only convey information about T in limited regions of the coordinate plane.

Free disposability of inputs FDI requires that

$$(x, y) \in T \Rightarrow (x', y) \in T$$

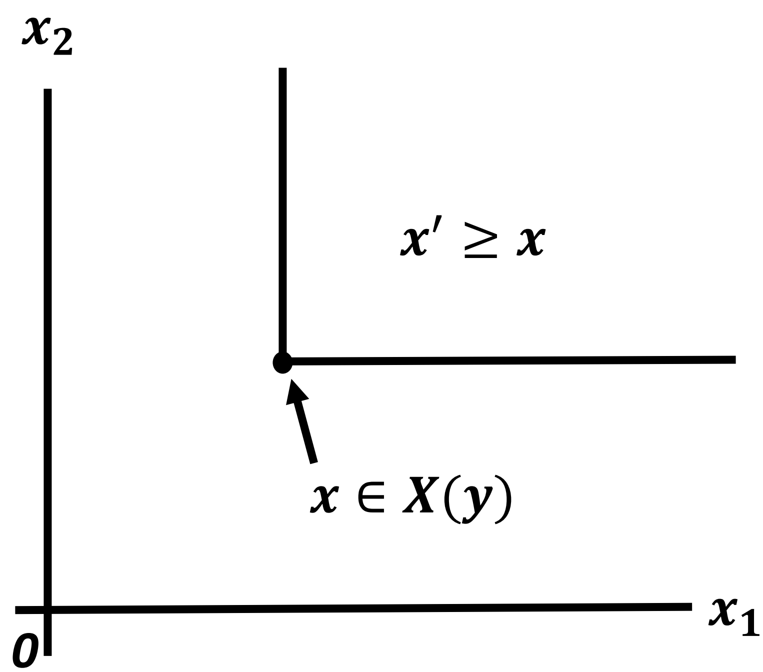
for $x' \geq x$. Using the equivalence relation $(x, y) \in T \Leftrightarrow x \in X(y)$, we rewrite this as:

- $FDI \Rightarrow\ x \in X(y) \Rightarrow x' \in X(y)$ for $x' \geq x$.

FDI requires that if x is in the input set for y, then any x' greater than or equal to x belongs to the input set for y.

We illustrate using Figure 3. Assume that the bundle illustrated by the point x belongs to $X(y)$. FDI then requires that any x' at least as large as x, depicted visually by any point to the northeast of x, also belongs to $X(y)$.

Figure 3: FDI and X(y)



Convexity Convexity means that T, when viewed as a set in (x, y) space, is convex. (In geometric terms, any two points that belong to T can be connected by a straight line segment that lies entirely within T.) When we consider input sets, we work in input, x, space with output, y, playing the role of an independent variate. In our discussion of convexity and

$X(y)$, we simplify by holding y constant. Keeping y fixed, the original definition of convexity becomes

$$(x^o, y) \in T \text{ and } (x', y) \in T \Rightarrow (\lambda x^o + (1-\lambda)x', y) \in T$$

for $\lambda \in (0,1)$. (Note that $\lambda y + (1-\lambda)y = y$.) Applying the equivalence relation $(x,y) \in T \Leftrightarrow x \in X(y)$ gives:

- *Convexity* $\Rightarrow x^o \in X(y)$ and $x' \in X(y) \Rightarrow \lambda x^o + (1-\lambda)x' \in X(y)$ for $\lambda \in (0,1)$.

We translate this into words as convexity implies that if x^o and x' belong to the input set for y, $X(y)$, then any weighted average of x^o and x' also belongs to $X(y)$. More simply, convexity implies that $X(y)$ is a convex set. This holds for all y.[2]

Figure 4 illustrates the consequences of imposing FDI and convexity on $X(y)$. We build the illustration as follows. First, assume that points x^o and x' belong to $X(y)$. Then impose convexity to obtain that any weighted average of these two points, illustrated by the line segment $x'x^o$ connecting them, also belongs to $X(y)$. Finally, use FDI to obtain that any points to the northeast of the points on the line segment $x'x^o$ also belong to $X(y)$.

Free disposability of output When there are multiple outputs, not all output bundles, y and y', can be ranked (ordered) using $\geq$. FDO only restricts production possibilities when two output bundles can be ranked using $\geq$. For those bundles that can be ranked, $y \geq y'$, FDO requires that $(x,y) \in T \Rightarrow (x,y') \in T$. Using the equivalence relation $(x,y) \in T \Leftrightarrow x \in X(y)$ gives

- *FDO* $\Rightarrow x \in X(y) \Rightarrow x \in X(y')$ for $y \geq y'$.

Or, in words, if x is in the input set for y, $X(y)$, it remains in the input set for y', $X(y')$, for for $y \geq y'$. Hence, every member of $X(y)$ must belong to $X(y')$, which we write in symbols as

- *FDO* $\Rightarrow X(y) \subset X(y')$ for $y \geq y'$.

FDO implies that the input set for y, $X(y)$, is a *subset* of $X(y')$ for $y \geq y'$. Figure 5 depicts this property as the boundary for $X(y)$ shifting towards the origin as y decreases to y'. The

[2]You might wonder what happens when $X(y) = \emptyset$. By convention, $\emptyset$ is convex.

Figure 4: FDI, Convexity, and X(y)

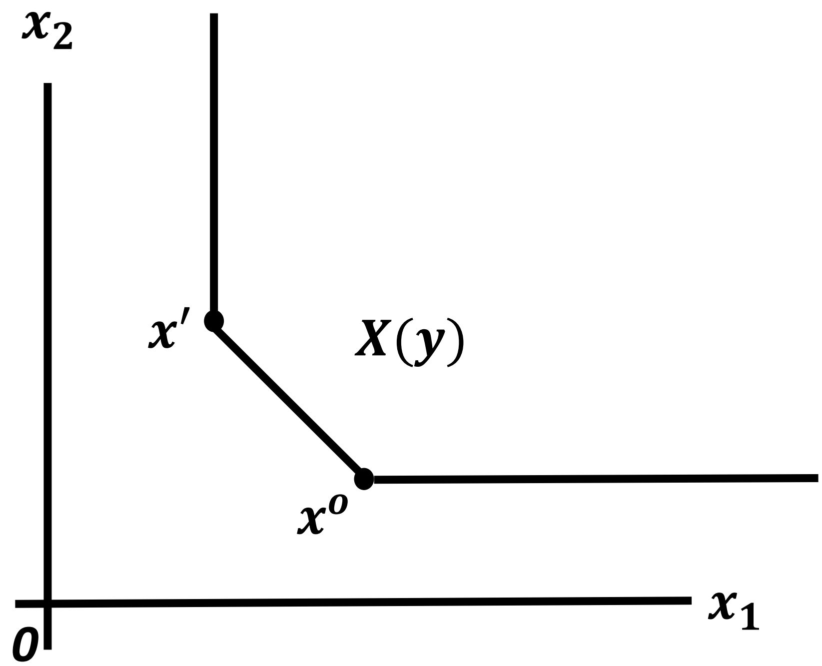
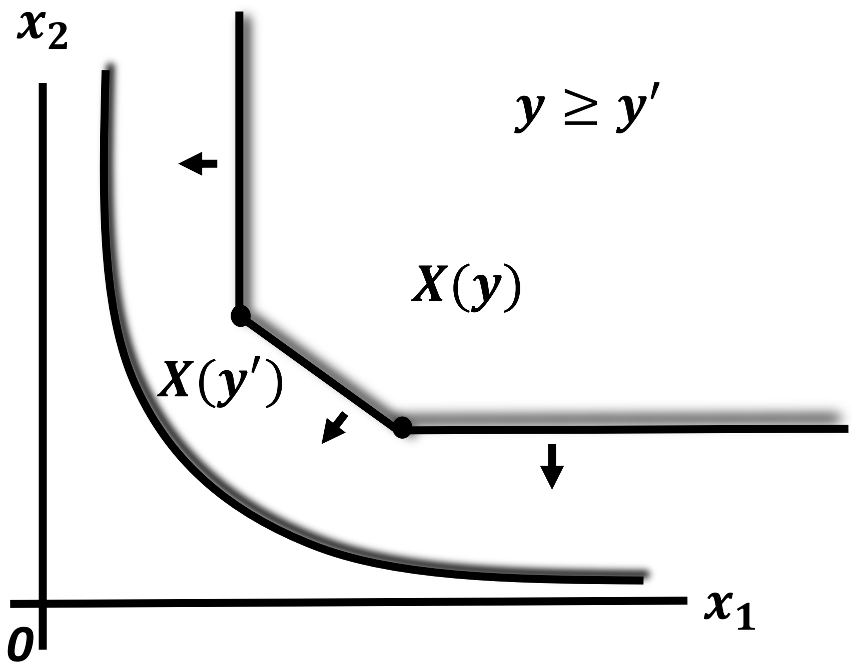


Figure 5: FDO and X(y)



input set '*expands*' (includes more elements) as y decreases. But FDO is also satisfied,if moving to y' from y leaves $X(y)$ unchanged, that is $X(y') = X(y)$. Strictly speaking, FDO rules out the input set getting smaller (having fewer elements) as output decreases.

No Free Lunch and No Fixed Costs NFL requires $(0, y) \notin T$ if $y \geq 0$ and $y \neq 0$ and NFC requires $(0, 0) \in T$. Using the equivalence relation $(x, y) \in T \Leftrightarrow x \in X(y)$, we write these as

- $NFL \Rightarrow 0 \notin X(y)$ if $y \geq 0$ and $y \neq 0$,
- $NFC \Rightarrow 0 \in X(0)$.

In visual terms, NFL says that the origin, 0, cannot belong to an input set for an output bundle, y, that contains a positive output. You cannot make something from nothing. NFC, on the other hand, says that you can use nothing to produce nothing.

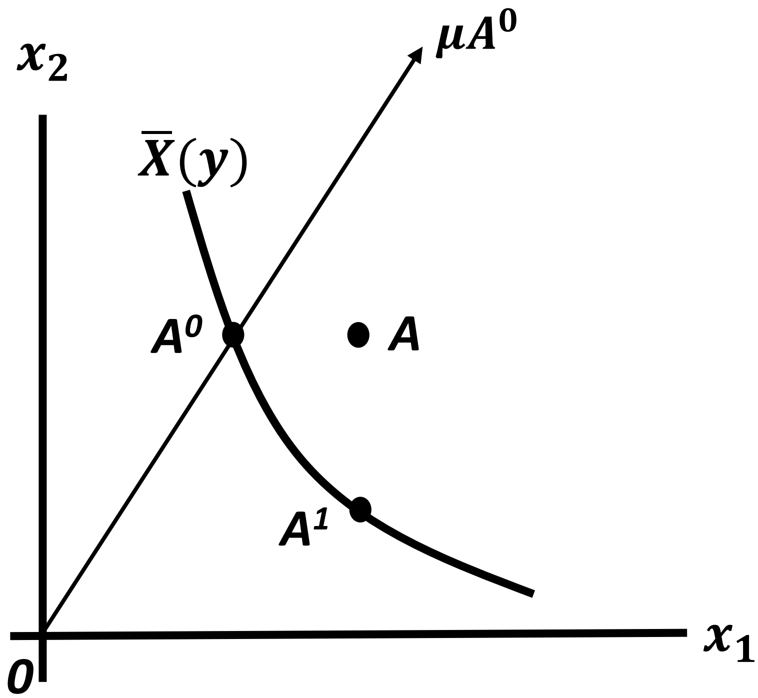
Isoquants and the Marginal Rate of Substitution

Figure 6 depicts the input-requirement set for a two-input canonical technology as everything on or above the curve labelled $\bar{X}(y)$. $\bar{X}(y)$ resembles the input-requirement function, $e(y)$, because it provides a lower boundary for the input bundles that can produce y, $X(y)$. But unlike $e(y)$, it contains a collection of points in the (x_1, x_2) plane rather than a single real number.

We call the curve $\bar{X}(y)$ the *isoquant* for y. Isoquant is a portmanteau word peculiar to economics. It combines *isos* from Greek meaning 'equal or like' with *quantity* from English to connote 'same quantity'.[3] Our use here captures the intuitive idea that all points on the curve represent input combinations that can produce the same y. Economists have a special interest in $\bar{X}(y)$ because it gives the set of input bundles from which self-interested producers will choose to produce y if they pay to purchase inputs. Consider, for example,

[3] Historians of economic thought debate to whom to assign credit for the notion of an isoquant. It parallels utility theory's indifference curve, which is attributed to F.Y. Edgeworth, 1881. *Mathematical Psychics: An Application of Mathematics to the Moral Sciences*. G.Kegan Paul. But it seems that the terminology originates with the Norwegian term 'isokvant' used in R. Frisch's lecture notes circa 1927.

Figure 6: An Isoquant for a Canonical Technology



point A in Figure 6. Because it's in $X(y)$, it can produce y. But so can A^0, which uses the same amount of x_2 but less of x_1. Similarly, A^1, uses the same amount of x_1 but less of x_2 than A. A self-interested individual who has to pay for inputs would choose either A^0 or A^1 over A. Points lying along curve segment A^0A^1 use less of both x_1 and x_2 than A to produce y.

Definition 2. An input vector, x, belongs to the *isoquant* for y, $\bar{X}(y)$, if and only if $x \in X(y)$ and any proportional shrinkage of x in the direction of the origin does not belong to $X(y)$.

Phrased in mathematical terms that definition becomes

$$\bar{X}(y) \equiv \{x \in X(y) : \lambda x \notin X(y) \texttt{ for } \lambda \in (0,1)\}.$$

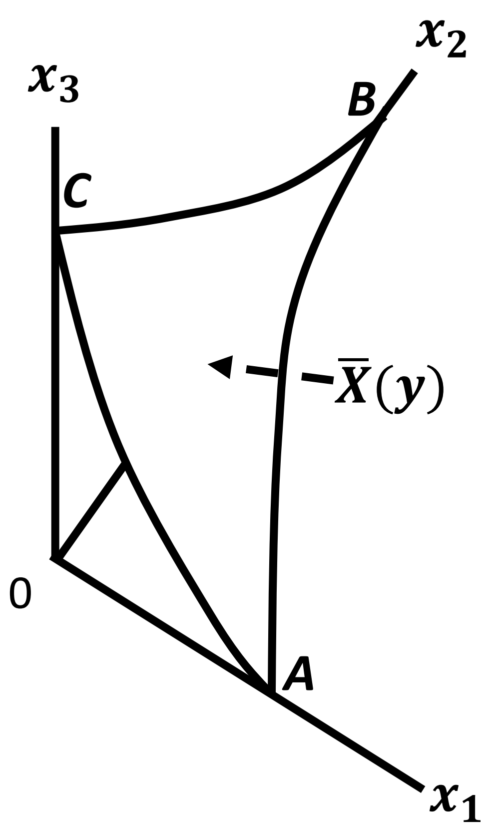
To visualize, take point A^0 in Figure 6. It belongs to $X(y)$. Consider the ray from the origin that passes through A^0. That ray is described in symbols as μA^0 where μ represents any positive real number, that is $\mu \in (0, \infty)$ $(0 < \mu < \infty)$. A^0 corresponds to $\mu = 1$. Points beyond A^0 on the ray correspond to $\mu > 1$ and points between the origin and A^0 correspond to $0 < \mu < 1$. Because A^0 lies in $X(y)$ and any movement along the ray towards the origin moves out of $X(y)$, it belongs to the isoquant. You can use a parallel argument to convince yourself that so does A^1. On the other hand, A can be moved towards the origin without taking it out of $X(y)$. Thus, it does not belong to $\bar{X}(y)$.

As depicted in Figure 6, the isoquant has a negative slope. That is a consequence of FDI, which rules out a positive slope. It also resembles an upturned bowl that faces towards the northeast. That characteristic is a consequence of convexity. Note that any two points on $\bar{X}(y)$ can be connected by a straight line segment that is contained in $X(y)$.

Figure 7 depicts an isoquant for a three-input canonical technology in (x_1, x_2, x_3) space as the manifold ABCA. You can confirm that the isoquant depicted in Figure 6 can be interpreted as fixing x_3 and then projecting the manifold ABC parallel to the x_3 axis onto the (x_1, x_2) plane. The same general principle applies for an arbitrary number of inputs.

A fundamental tenet of economics is that self-interested individuals respond to changing or evolving market conditions by altering their economic activities. Therefore, if prices

Figure 7: Isoquant for Three Input Technology



change, producers respond by adjusting their behavior to accommodate those price changes. In particular, if certain inputs become more expensive, producers will find their use less attractive and may attempt to find more attractive input combinations to produce y. Economists refer to such behavior as producers *substituting* some inputs for others.

To get a handle on what's involved, let's assume that the producer is at A^0 in Figure 6. If that input bundle becomes unattractive, does the producer have alternative ways to produce y? The answer is yes, any input bundle in $X(y)$ will also produce y. And, of those alternatives, the ones on $\bar{X}(y)$ are the most attractive. So, suppose that the producer moves along the isoquant to A^1 from A^0. Making that move involves the producer using less of x_2 and more of x_1, while holding y constant. Speaking economics, x_1 *has been substituted for* x_2 in the production of y

To measure the degree to which inputs substitute for one another economists use:

Definition 3. The *marginal rate of substitution* (MRS) of input x_1 for input x_2 is the rate at which marginal changes in x_1 are substituted in $\bar{X}(y)$ for marginal changes in input x_2 while holding output y and all other inputs constant.

Before we offer a visual interpretation of this definition, a few comments are in order. First, inputs 1 and 2 act as placeholders here. For a multiple-input technology, one can define a MRS between, say, inputs k and j by replacing x_1 and x_2 with, respectively, x_k and x_j. Second, when we speak of a MRS, we are talking about how two, *and only two*, inputs interact as all other inputs and output are held constant. Third, 'rate' when used in economic discussions always signals a ratio comparison. And fourth, under FDI, the MRS can never be positive. It can be 0 or even $-\infty$, but not positive.[4]

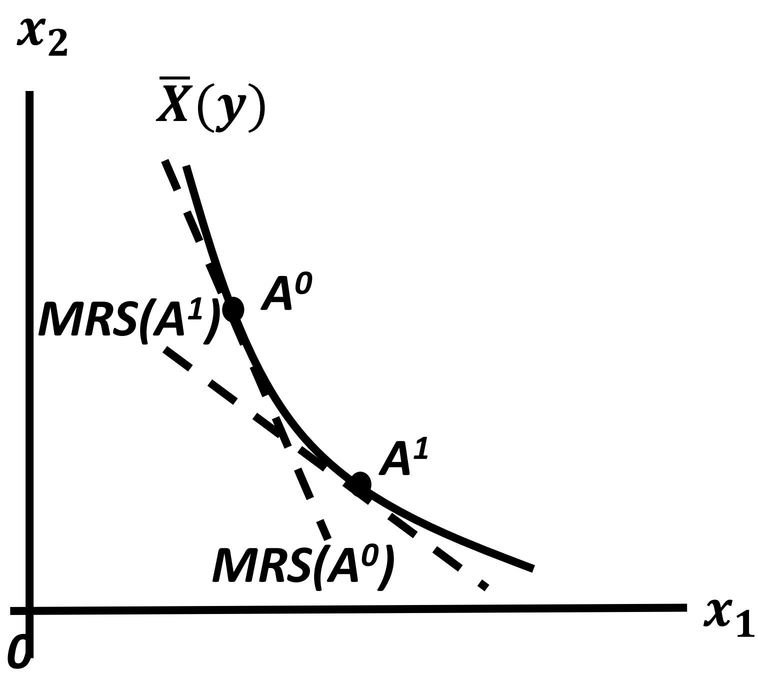
Presuming, as in our discussion of marginal products, that marginal is interpreted as involving 'tiny' changes in x_1 and x_2 gives the visual definition for the MRS:

Definition 4. The marginal rate of substitution at x is given by the slope of the line segment tangent to the isoquant at x.

[4]Some (perhaps most) authors define the MRS as the negative of our definition to avoid dealing with negative numbers. To preserve simplicity and to avoid confusion, we do not follow that convention.

Figure 8 illustrates with the MRS at A^0 given by the slope of the dotted line segment $MRS\left(A^0\right)$ and the MRS at A^1 given by the slope of the dotted line segment $MRS\left(A^1\right)$. As required under FDI both slopes are negative, and (when viewed from the perspective of

Figure 8: Diminishing Marginal Rate of Substitution



the x_2 axis) $MRS\left(A^1\right)$ has a flatter slope than $MRS\left(A^0\right)$. This flattening of the isoquant is referred to as the technology exhibiting a *diminishing marginal rate of substitution* and is yet another consequence of imposing convexity upon T.

To interpret this phenomenon, consider a thought experiment that involves the producer decreasing x_2 successively by the same amount while replacing the lost production potential by increasing x_1. Viewed this way, the MRS measures the change in x_1 needed to balance the x_2 decrease. For each successive increment that the producer decreases x_2, the principle of a diminishing marginal product suggests each matching decrease in y is larger. (See, for example, the single-output technology in Figure 1). But that same principle means that larger and larger doses of x_1 are needed to balance the losses in y. So, x_1 becomes a poorer and poorer substitute for x_2 as you move from, say, A^0 to the southeast along $\bar{V}(y)$ in Figure 8. The visual manifestation of becoming a poorer substitute is a flattening of the isoquant. If you reverse the thought experiment and reason in terms of decreasing x_1 while increasing x_2 (moving to the northwest from A^0), a parallel argument shows that larger and larger incremental doses of x_2 will be needed to balance successive decreases in x_1 of the same size.

Just as the notion of a marginal product presents difficulties when the graph of the production function is kinked, so too does the MRS notion when the isoquant is kinked. Take, for example, Figure 4. Working from the northwest to the southeast, the isoquant for that $X(y)$ consists of the vertical line segment that emanates from x', the line segment connecting x' and x^o, and the horizontal line segment emanating from x^o. Along the vertical line segment, x_2 can be decreased without requiring any balancing adjustment in x_1. The rate at which x_1 substitutes for a tiny decrease in x_2 of the amount φ is $\frac{-\varphi}{0}$, which equals $-\infty$ by convention. The portion of $\bar{X}(y)$ lying on the line segment $x'x^o$ has a negative (but finite) slope, and the horizontal portion has slope 0. Thus, any line segment with slope between $-\infty$ and the slope of the line segment $x'x^o$ will be tangent to the isoquant at x'. Similarly, any line segment with slope between the slope of the line segment $x'x^o$ and 0 will be tangent at x^o. Kinks in isoquants are associated with multiple (an infinity) of different marginal rates of substitution.[5] Kinks in isoquants have important implications for producer behavior. (Lecture 6 treats this issue).

Output Sets

Definition 5. The *output set* for x, $Y(x)$, gives the collection of output bundles that can be produced by input bundle x

[5]The set of real numbers falling between the slope of $x'x^o$ and 0 has infinitely many members.

Written in mathematical terms, that becomes

$$\begin{aligned} Y(x) &\equiv \{y : x \text{ can produce } y\} \\ &= \{y : (x, y) \in T\} \\ &= \{y : x \in X(y)\} \end{aligned}$$

Under free disposability of outputs,

$$Y(x) = \{y : f(x) \geq y\},$$

for a single-output technology so that the production function is the upper bound of $Y(x)$. Thus, $Y(x)$ resembles the production function in that it takes x as the independent variable and defines a range of producible output but different because it isolates *sets of output vectors* rather than a single output.

The following equivalence relations are immediate consequences of the definitions:

$$x \in X(y) \Leftrightarrow (x, y) \in T \Leftrightarrow y \in Y(x).$$

These relations reveal that in each case we study the same object, the technology, but from different perspectives. The equivalence relation $(x, y) \in T \Leftrightarrow y \in Y(x)$ is crucial in establishing the properties of $Y(x)$.

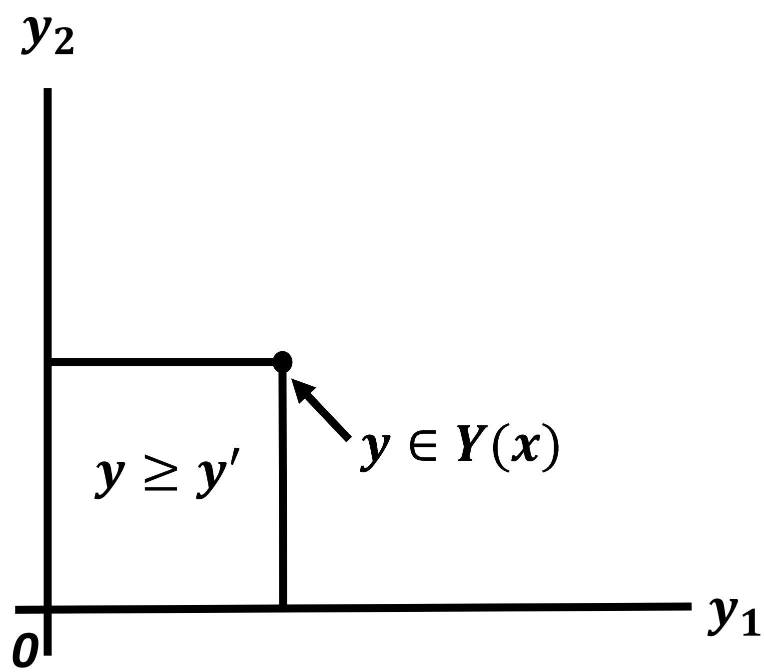
T not empty If $T \neq \emptyset$, a feasible (x, y) bundle must exist. Take that x and note that by definition $y \in Y(x)$. That gives:

- $T \neq \emptyset \Rightarrow \exists\, x : Y(x) \neq \emptyset$.

T not empty ensures that at least one input bundle, x, has a nonempty output set, $Y(x)$. It does not imply that $Y(x) \neq \emptyset$ for all x.

Free Disposability of Output FDO requires that if x can produce y, it can produce any output bundle y' whose elements are all smaller than the corresponding element in y. Figure 9 illustrates by choosing a point y that belongs to the output set for x. The output bundles, y' that satisfy $y \geq y'$ are those to the southwest of y. FDO requires that all belong to $Y(x)$. Applying the equivalence relationship, $(x, y) \in T \Leftrightarrow y \in Y(x)$, to

Figure 9: FDO and Y(x)



- $FDO \Rightarrow (x, y) \in T$ then $(x, y') \in T$ for $y \geq y'$

gives

- $FDO \Rightarrow y \in Y(x)$ then $y' \in Y(x)$ for $y \geq y'$.

Convexity We have already learnt that convexity requires T to be a convex set in (x, y) space and $X(y)$ to be a convex set in x space. It's reasonable to conjecture that convexity requires $Y(x)$ to be a convex set in y space. Showing that the conjecture is true is an easy consequence of the definition of convexity and the equivalence relation, $(x, y) \in T \Leftrightarrow y \in Y(x)$. The definition of convexity requires

$$(x^o, y^o) \in T \;\; \texttt{and} \;\; (x', y') \in T \Rightarrow \lambda (x^o, y^o) + (1 - \lambda)(x', y') \in T$$

for $\lambda \in (0, 1)$. Now take $x^o = x^{'} = x$ so that we keep the input constant. Then using the equivalence relation gives:

- $Convexity \Rightarrow y^o \in Y(x) \;\; \texttt{and} \;\; y' \in Y(x) \Rightarrow \lambda y^o + (1 - \lambda) y^{'} \in Y(x)$

for $\lambda \in (0, 1)$. All weighted averages of y^o and y' belonging to $Y(x)$ also belong to $Y(x)$. $Y(x)$ is a *convex* set in x space.

You can visualize the impact of imposing convexity by plotting two points y^o and y' in the (y_1, y_2) plane. (The actual artwork is left to you as part of Exercise 4.) Convexity implies that $Y(x)$ contains all of the points on a line segment connecting the two points.

Free disposability of inputs Recall from Lecture 3 that FDI implies that $f(x)$ is nondecreasing in x (nonnegative marginal product). Intuition, therefore, suggests that a similar property applies here. And it does.

The argument is as follows. Isolate $Y(x)$ for x. Now consider what happens when the input bundle is increased in any of its N dimensions to x'. FDI requires that any output bundle that could be produced using x can still be produced. That means that every point in $Y(x)$ must also belong to $Y(x')$ for $x' \geq x$. $Y(x)$ is a subset of $Y(x')$, that is

- $FDO \Rightarrow \; Y(x) \subset Y(x')$ for $x' \geq x$

Output sets do not get smaller as the input bundle grows. Or, more output bundles can be produced using larger input bundles than smaller ones. (You are to illustrate the result in Exercise 5.)

Three points to consider. First, you can arrive at this conclusion by using the original definition of FDI in terms of T, applying the equivalence relationship, $(x, y) \in T \Leftrightarrow y \in Y(x)$, and then using the definition of a subset. Second, FDI is a statement about x' and x that can be ranked using $\geq$. Not all input bundles can be ranked according to $\geq$ when there are multiple inputs. FDI imposes no restrictions over input bundles that cannot be ranked using $\geq$. Third, we speak in terms of not getting smaller to ensure precision. But the strong intuition is that increasing input bundles expands output sets.

No Free Lunch and No Fixed Cost Deriving the consequences of NFL and NFC for $Y(x)$ is a simple matter of applying the equivalence relationship, $(x, y) \in T \Leftrightarrow y \in Y(x)$, to

- $NFL \Rightarrow (0, y) \notin T$ if $y \geq 0$ and $y \neq 0$ and $NFC \Rightarrow (0, 0) \in T$,

to obtain

- $NFL \Rightarrow y \notin Y(0)$ if $y \geq 0$ and $y \neq 0$ and $NFC \Rightarrow 0 \in Y(0)$.

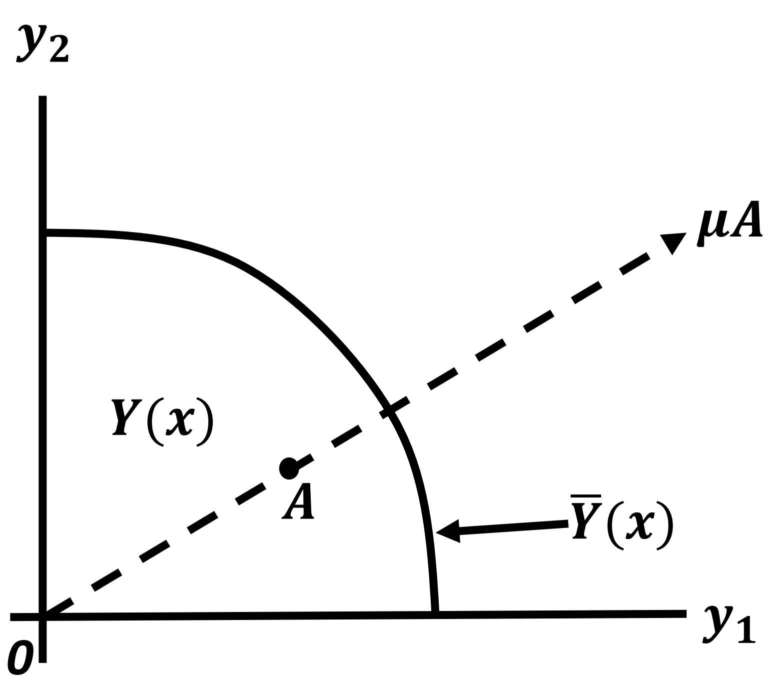
In words, zero input cannot produce a positive input, and zero input can produce zero output.

Transformation Curve and the Marginal Rate of Transformation

Figure 10 depicts the output set for x as everything on or below the curve labelled $\bar{Y}(x)$. As drawn, $Y(x)$ is consistent with a two-output, multiple-input canonical technology. You can verify for yourself that it satisfies FDO and convexity. FDI will require that $Y(x)$ gets no smaller as x expands, and NFL is satisfied so long as $x \neq 0$.

We have isolated a point A, through which we have drawn the ray μA. That ray consists of all *radial* (that is along a ray) expansions or contractions of output bundle A. If you

Figure 10: The Transformation Curve



compare A with the point where μA intersects the curve $\bar{Y}(x)$, it's clear that the latter has more of both y_1 and y_2 than the former. Therefore, a self-interested individual who owned the rights to selling the output bundles produced using x would prefer the point on $\bar{Y}(x)$ to A. If you apply this argument to all points lying below the curve $\bar{Y}(x)$ you arrive at a similar conclusion.

The curve $\bar{Y}(x)$ has several aliases in economics. We call it the *transformation curve for* x.[6] But you may see it called, among others, the *production possibilities curve*, the *production possibilities frontier*, or the *transformation frontier*. Regardless of semantics, the set of output bundles it portrays is of special interest to economists. It gives the set of output bundles in $Y(x)$ at which self-interested individuals will locate their production choices. The analogy with the isoquant is exact. To produce a given output, self-interested producers will locate on the isoquant. In employing a fixed input bundle, self-interested producers will locate on the transformation curve. If they do not, they forego the ability to realize revenue from their production.

$\bar{Y}(x)$'s definition parallels that of the isoquant.

Definition 6. An output vector, y, belongs to the *transformation curve* for x, $\bar{Y}(x)$, if and only if y belongs to the output set for x but any proportional expansion of y outward from the origin does not belong to the output set for x

The corresponding mathematical statement is

$$\bar{Y}(x) \equiv \{y \in Y(x) : \lambda y \notin Y(x) \quad \text{for } \lambda > 1\}.$$

Thus, $\bar{Y}(x)$ provides an *outer boundary* for $Y(x)$. Any radial movement from it to the northeast carries the output bundle out of the output set.

Just as we are interested in how inputs can be substituted for one another as economic circumstances change, we are also interested in how producers, using a fixed input bundle, respond to changing economic circumstances. Because self-interested producers will locate on the transformation curve, economists have developed a notion of adjustments along that curve that parallels the idea of a marginal rate of substitution. We have:

[6] Given our definition, "transformation set" would be more precise but that clashes with more familiar economic jargon.

Definition 7. The *marginal rate of transformation* (MRT) of output y_1 for output y_2 is the rate at which marginal changes in y_1 in $\bar{Y}(x)$ are transformed into marginal changes in output y_2 while holding input x and all other outputs constant.

The notion of *transformation* that is invoked is that of producers changing the mix of their output bundles as they move along the boundary of $\bar{Y}(x)$. Starting from the point where $\bar{Y}(x)$ intersects the horizontal axis so that a positive amount of y_1 but no y_2 is produced, y_1 is *transformed* into y_2 as one moves to the northwest along $\bar{Y}(x)$. As in the case of the marginal rate of substitution, y_1 and y_2 are placeholders here for arbitrary outputs. It makes sense to speak of a marginal rate of transformation between any two outputs holding the other outputs and inputs constant. Our geometric definition is:

Definition 8. Marginal Rate of Transformation (MRT) at $y \in \bar{Y}(x)$ equals the slope of transformation curve at y.

Referring to Figure 10, you will see that the MRT is nonpositive for all points on $\bar{Y}(x)$.[7] It can approach zero, but it will never become positive. This is a consequence of FDO. If FDO is violated, we cannot argue that the slope remains nonpositive. Moreover, as you move from the point of intersection between $\bar{Y}(x)$ and the vertical axis to the southeast along the transformation curve, its slope becomes steeper (more nearly vertical). That is a consequence of convexity and is referred to as an *increasing marginal rate of transformation.*

This phenomenon is the 'flip-side' of the 'decreasing rate of marginal substitution' and is illustrated using a parallel thought experiment. Imagine that you are a producer with a fixed bundle of inputs. You can choose how to allocate those inputs to different activities, but you cannot vary their absolute amount. Now starting at a point where all of your x is devoted to producing y_2 (where $\bar{Y}(x)$ intersects the vertical axis) imagine trying to increase the production of y_1 successively by the same amount. As we showed in the single-output case with $e(y)$, this requires larger and larger reallocations of that fixed input bundle towards producing y_1. But diverting resources from y_2, by the principle of diminishing marginal

[7] Many authors define the marginal rate of transformation in absolute value terms. Just as we avoided defining the marginal rate of substitution in absolute value terms, we avoid the similar convention for MRT.

returns, evokes ever larger decreases in y_2. The visual manifestation is a steeper slope for $\bar{Y}(x)$.

A Closing Word About Convexity*

Imposing convexity upon T ensures that the graphs of $f(x)$, $e(y)$, $\bar{X}(y)$, and $\bar{Y}(x)$ possess a curvature that is consistent with the generalized principle of diminishing returns. From our presentation, it's natural to construe that a diminishing marginal productivity, a diminishing marginal rate of substitution, and an increasing marginal rate of transformation all manifest the same phenomenon.

While intuitive, it's not correct. Our discussion shows that imposing convexity on T implies a diminishing marginal product, a diminishing rate of substitution, and an increasing marginal rate of transformation. But it is not true that if $X(y)$ is a convex set, which is the basis for the diminishing marginal rate of substitution, that T must be a convex set. Nor is it true that if $Y(x)$ is a convex set that T must be convex. Thus, while a diminishing marginal rate of substitution and an increasing rate of transformation manifest phenomena similar to a diminishing marginal product they are different.

To illustrate, we use the equivalence relationships

$$x \in X(y) \Leftrightarrow (x, y) \in T \Leftrightarrow y \in Y(x).$$

For $X(y)$ to be a convex set, all weighted averages (convex combinations) of all its elements must belong to $X(y)$. We write this as

$$x^o \in X(y) \ \texttt{and} \ x' \in X(y) \Rightarrow \lambda x^o + (1-\lambda) x' \in X(y) \ \texttt{for} \ \lambda \in (0,1).$$

Using the equivalence relationship shows that this requires that

$$y \in Y(x^o) \ \texttt{and} \ y \in Y(x') \Rightarrow y \in Y(\lambda x^o + (1-\lambda) x') \ \texttt{for} \ \lambda \in (0,1). \tag{1}$$

Expression (**??**) requires that if y belongs to both $Y(x^o)$ and $Y(x')$, it must belong to the output sets for all weighted averages of x^o and x'. But if y belongs to $Y(x^o)$ and $Y(x')$, then it belongs to their intersection, $y \in Y(x^o) \cap Y(x')$. Thus, expression (**??**) becomes

$$y \in Y(x^o) \cap Y(x') \Rightarrow y \in Y(\lambda x^o + (1-\lambda) x') \text{ for} \lambda \in (0,1). \tag{2}$$

Expression (**??**) does not require that $Y(x)$ is convex as a set, which is required by convexity of T.[8] Therefore, convexity of $X(y)$ and T as sets have different consequences for $Y(x)$. They are different properties. And because convexity of T implies convexity of $X(y)$, but not the other way around, convexity of T is the *stronger* restriction.

[8]For the single-output, multiple-input technology technology, this condition requires that $f(x)$ is *quasi-concave*, which is a weaker condition than concave.

Exercises

1. Explain why nonemptines of T $(T \neq \emptyset)$ does not imply that $X(y) \neq \emptyset$ for all y.

2. Suppose that you know that $x' \in X(y)$ and $x^o \in X(y)$. Impose free disposability of inputs (FDI) and illustrate the result. Impose convexity on the resulting set and illustrate the result. Compare and contrast it with Figure 4.

3. Explain why nonemptines of T $(T \neq \emptyset)$ does not imply that $Y(x) \neq \emptyset$ for all x.

4. Suppose that you know that $y' \in Y(x)$ and $y^o \in Y(x)$. Impose convexity T and illustrate the consequences for $X(y)$. Illustrate what happens if you then impose free disposability of output (FDO).

5. Illustrate and discuss the consequence of imposing FDI upon $Y(x)$.

6*. Illustrate the property described in (**??**) geometrically.

7*. Derive and illustrate geometrically the consequences of requiring $Y(x)$ to be a convex set (without imposing convexity on T) for $X(y)$.

Lectures on Neoclassical Production Economics[1]

Lecture 5: A Primer on Rational Producers

Robert G Chambers[2]

January 9, 2024

[2]University of Maryland

Lecture Introduction

So far, our focus has been on the technology, T. In this lecture, we start our economic analysis by introducing the concepts of *profit*, *cost*, *revenue*, and their geometric illustrations. We then use a simple single-input, single-output model to study the basic principles of profit maximization. We close by showing that profit-maximizing producers always minimize cost and maximize revenue. That demonstration identifies the relationships between three important economic objects, the *profit function*, the *cost function*, and the *revenue function* that provide the primary focus of Lectures 6 through 9.

Profit, Cost, and Revenue: Definitions and a Geometric Overview

As explained in Lecture 2, we treat producers who operate in competitive markets and face fixed prices that they cannot manipulate. The prices for inputs are strictly positive, and we denote them by the N-dimensional real vector $w = (w_1, w_2, ..., w_N)$. Here, for example, w_n represents the price of the nth input. The prices at which producers sell their outputs are strictly positive and are denoted by the M-dimensional real vector $p = (p_1, p_2, ..., p_M)$.

A self-interested, price-taking producer seeks the largest possible *profit*. *Profit* is the difference between the producer's *revenue* and the producer's *cost*. We denote it by π (the Greek letter 'pi') so that

$$\pi \equiv r - c,$$

where

$$r \equiv py \equiv \sum_m p_m y_m = p_1 y_1 + p_2 y_2 + ... + p_M y_m,$$

the sum taken over all outputs of each output price times the amount sold, is *revenue*, and

$$c \equiv wx \equiv \sum_n w_n x_n = w_1 x_1 + w_2 x_2 + ... + w_N x_N,$$

the sum taken over all inputs of the input price times the amount purchased, is *cost*.

The assumption that prices are positive ($p > 0$ and $w > 0$) means that markets evaluate inputs and outputs as "goods" rather than as "bads". We saw in Lecture 2 that inputs and outputs can be undesireable, but our basic producer model abstracts from such issues to narrow the focus on the economic principles guiding producer behavior.[1] Because we treat profit maximizers, we assume that the producer has the exclusive right to revenue from output sales and the sole responsibility for production costs.

Prices, profit, cost, and revenue are expressed in currency units. Economists call prices, profit, cost, or revenue denominated in currency units (be they dollars, euros, renminbi, etc.) *nominal.* In contrast, the *real* or *relative* price of commodity A in units of commodity B is derived by dividing A's nominal price by B's nominal price to obtain p_A/p_B.[2] When B's price is used as the divisor in creating a real price, B is called the *numeraire.* The real price of the numeraire is thus 1. Real profit, cost, and revenue are derived in a similar fashion.

Economists often distinguish nominal changes from real changes by representing a nominal change as multiplying the nominal value of each variable by a common positive real number, $\mu > 0$. Such changes are also called *proportional* because the percentage changes are common across the variables. For example, by moving from p to $3p$ (so that $\mu = 3$), the percentage change in the mth price is

$$\frac{3p_m - p_m}{p_m} = 2$$

which is common for all m. Such changes are nominal because they leave relative prices unchanged. For example, if output 1 is the numeraire then before and after multiplication of nominal prices by a common positive real number the real prices satisfy

$$\left(1, \frac{p_2}{p_1}, ..., \frac{p_M}{p_1}\right) = \left(1, \frac{\mu p_2}{\mu p_1}, ..., \frac{\mu p_M}{\mu p_1}\right).$$

Visual illustrations of profit, revenue, and cost are key to explaining our ideas. Take profit, and assume that T has a single input and a single output. Profit is now a linear function of the input and the output (x, y),

$$\pi = py - wx.$$

We first rewrite this expression with output, y, treated as the dependent variable of a function of input, x, and *real profit*, π/p. Divide both sides of the profit definition by the output price,

[1]The appropriate modelling of good and bad inputs and outputs remains controversial, especially in areas of economics, such as environmental economics, that often ignore technical production issues in their models.

[2]The terms, *real* and *relative*, are used interchangeably throughout the lectures.

p, and add $\frac{w}{p}x$ to both sides to get

$$y = \frac{\pi}{p} + \frac{w}{p}x.$$

Figure 1 depicts this linear function as the straight line that intersects the vertical axis (measuring y) at π/p (real profit in units of output) that has slope in x of w/p (real price of the input in units of the output).[3] This line traces out the set of points in the (x, y) plane that give a real profit of π/p when prices are (w, p). We call it the *isoprofit line (hyperplane)* for (π, p, w), and represent it in symbols as $\bar{\mathcal{P}}(\pi, p, w)$. (A hyperplane is a higher dimensional version of a line.)

As its name suggests, the isoprofit line is similar to the isoquant (and the transformation curve) because it isolates a collection (set) of quantities that give the same outcome. It differs, however, because it arises from producer interaction with competitive markets and not the technology, T. Its formal definition (which applies regardless of the dimension of x and y) is:

Definition 1. An *isoprofit line* (hyperplane) for (π, p, w), $\bar{\mathcal{P}}(\pi, p, w)$, is the set of input-output bundles that give a profit of π at prices (p, w).

We write this verbal definition in mathematical symbols as:

$$\bar{\mathcal{P}}(\pi, p, w) \equiv \{(x, y) : \ \pi = py - wx\}.$$

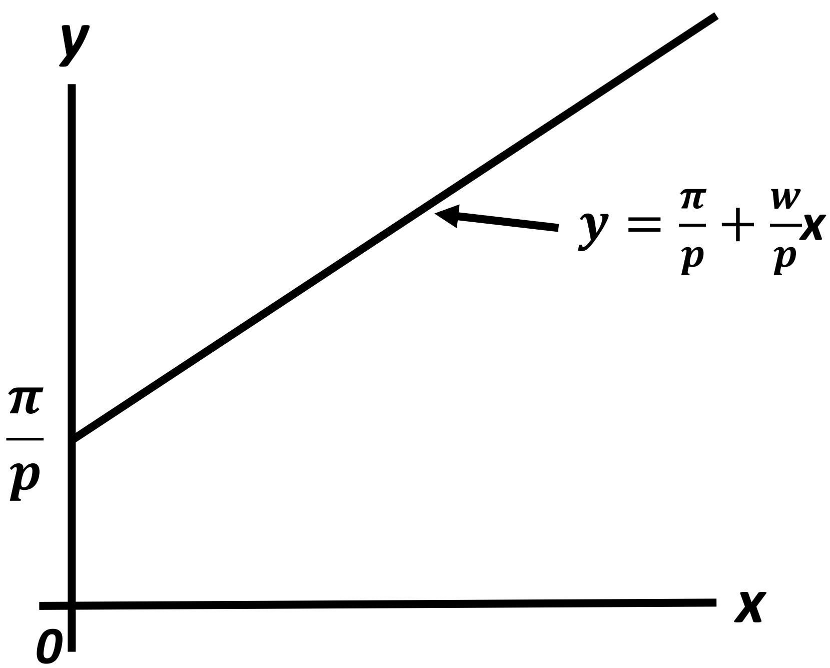
Exercise 1 asks you to show that: if (π, p, w) vary proportionately so that $(\pi/p, w/p)$ remain unchanged, $\bar{\mathcal{P}}(\pi, p, w)$ remains unchanged (the isoprofit line is *homogeneous of degree zero* in (π, p, w)); if π increases holding (p, w) constant, the vertical intercept of the isoprofit line shifts up, and the slope remains unchanged resulting in a parallel upwards shift; if w increases while holding (π, p) constant, so that inputs become more expensive, the slope of

[3] The choice of the dependent variable and the independent variable in Figure 1 is arbitrary. We use y because our convention is to represent T with outputs depicted on the vertical axis and inputs on the horizontal axis. But if you choose x to be the dependent variable, you obtain

$$x = \frac{p}{w}y - \frac{\pi}{w},$$

as the formula for the isoprofit line. If you graph this version in (x, y) space you obtain Figure 1.

Figure 1: Isoprofit Line $\bar{\mathcal{P}}(\pi, p, w)$



the isoprofit line becomes steeper without changing the vertical intercept; and if p increases while holding (π, w) constant, the vertical intercept shifts downward and the slope becomes flatter. (You, of course, are encouraged to also consider what happens when these increases are replaced by decreases.)

The areas in (x, y) space on or above an isoprofit line (hyperplane)

$$\mathcal{P}^{+}(\pi, p, w) \equiv \{(x, y) : py - wx \geq \pi\}$$

or on or below it

$$\mathcal{P}^{-}(\pi, p, w) \equiv \{(x, y) : py - wx \leq \pi\}$$

are called closed *half spaces* in math jargon. Because they are defined as everything on or to one side of a line (hyperplane), half spaces form convex sets in (x, y) space. You can verify this by inspecting Figure 2. For example, take any two points falling on or below the isoprofit line $\bar{\mathcal{P}}(\pi, p, w)$ in the area labelled $\mathcal{P}^{-}(\pi, p, w)$. If you connect them with a straight line segment, all the resulting weighted averages fall on or below the isoprofit line.

Turning to cost, let N=2 so that cost, c, is a linear function of x_1 and x_2

$$c = w_1 x_1 + w_2 x_2.$$

To represent this in the (x_1, x_2) plane, we rewrite it as

$$x_2 = \frac{c}{w_2} - \frac{w_1}{w_2} x_1$$

and depict it in Figure 3 by the line with negative slope $-\frac{w_1}{w_2}$, vertical intercept $\frac{c}{w_2}$, and horizontal intercept $\frac{c}{w_1}$. It gives all the points in the (x_1, x_2) plane that incur a cost of c when input prices are (w_1, w_2). We call this line the *isocost line* (hyperplane) and denote it by $\bar{\mathcal{C}}(c, w)$. Its definition (which applies regardless of the dimension of x) is

Definition 2. An *isocost line* (hyperplane) for (c, w), $\bar{\mathcal{C}}(c, w)$, is the set of input bundles that give a cost of c at input prices w.

The mathematical version is:

$$\bar{\mathcal{C}}(c, w) \equiv \{x : c = wx\}$$

Figure 2: Profit Half Spaces

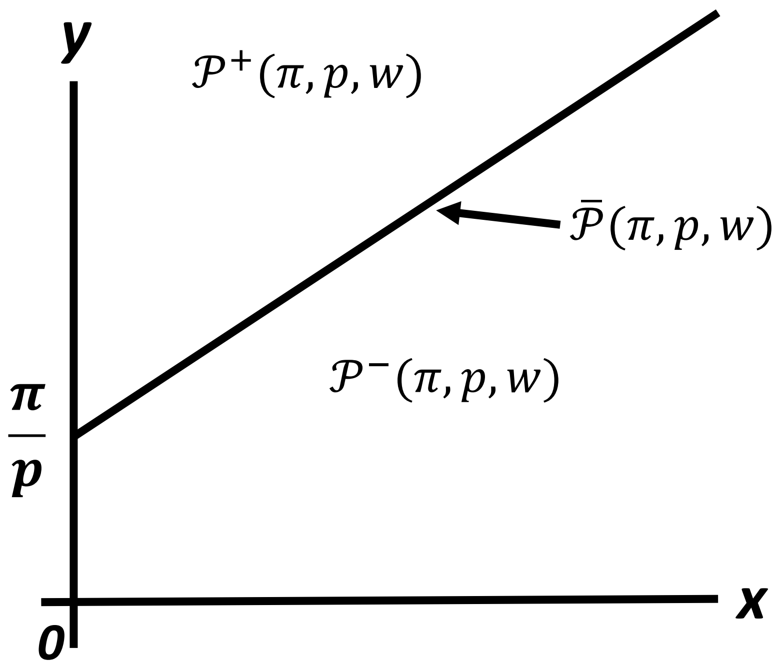
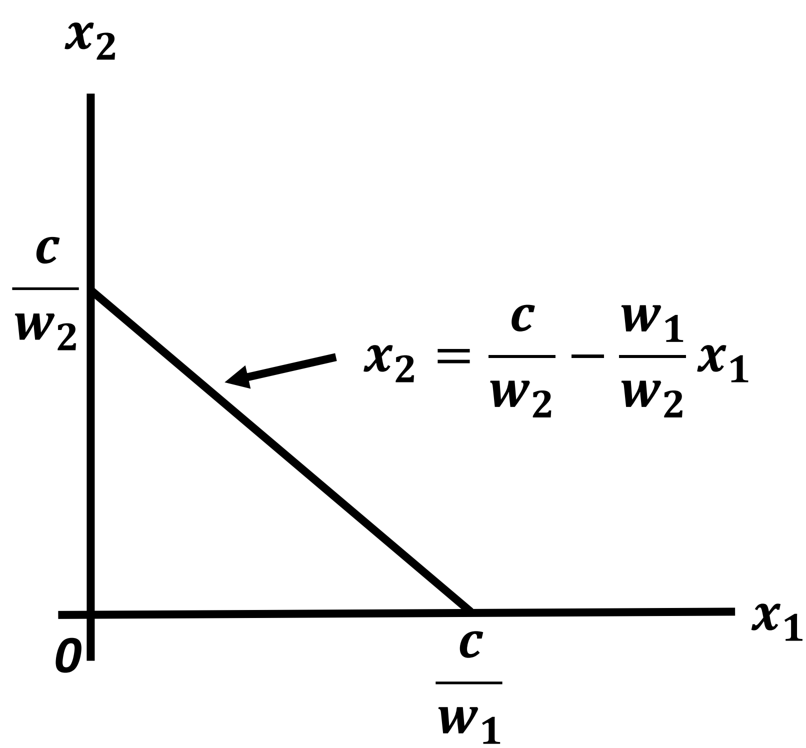


Figure 3: Isocost Line $\bar{\mathcal{C}}(c, w)$



Exercise 2 asks you to show that: if (c, w_1, w_2) vary proportionately keeping $(c/w_2, w_1/w_2)$ constant, $\bar{\mathcal{C}}(c, w)$ is unchanged ($\bar{\mathcal{C}}(c, w)$ is *homogeneous of degree zero* in (c, w)); decreasing c while leaving (w_1, w_2) unchanged causes a *parallel shift* in the isocost line towards the origin; increasing w_1 holding c and w_2 constant causes the horizontal intercept of $\bar{\mathcal{C}}(c, w)$ to shift towards the origin and the slope of the isocost line to become steeper (when viewed from the perspective of x_2); increasing w_2 while holding c and w_1 constant causes the horizontal intercept to shift towards the origin and the slope of the isocost cost line to become flatter.

The half spaces in (x_1, x_2) space associated with $\bar{\mathcal{C}}(c, w)$ are:

$$\mathcal{C}^+(c, w) \equiv \{x : wx \geq c\}$$

and

$$\mathcal{C}^-(c, w) \equiv \{x : wx \leq c\}.$$

That brings us to revenue. Let M=2 (two outputs) so that

$$r = p_1 y_1 + p_2 y_2.$$

Mimicking arguments made for π and c, we transform this expression into a function of real revenue ($\frac{r}{p_2}$) and real revenue from sales of good 1 ($\frac{p_1}{p_2} y_1$) with y_2 as the dependent variable

$$y_2 = \frac{r}{p_2} - \frac{p_1}{p_2} y_1$$

that we illustrate in Figure 4 by the line segment with vertical intercept $\frac{r}{p_2}$, slope $-\frac{p_1}{p_2}$, and horizontal intercept $\frac{r}{p_1}$. We refer to that line as the *isorevenue line*, denote it by $\bar{\mathcal{R}}(r, p)$, and define it as

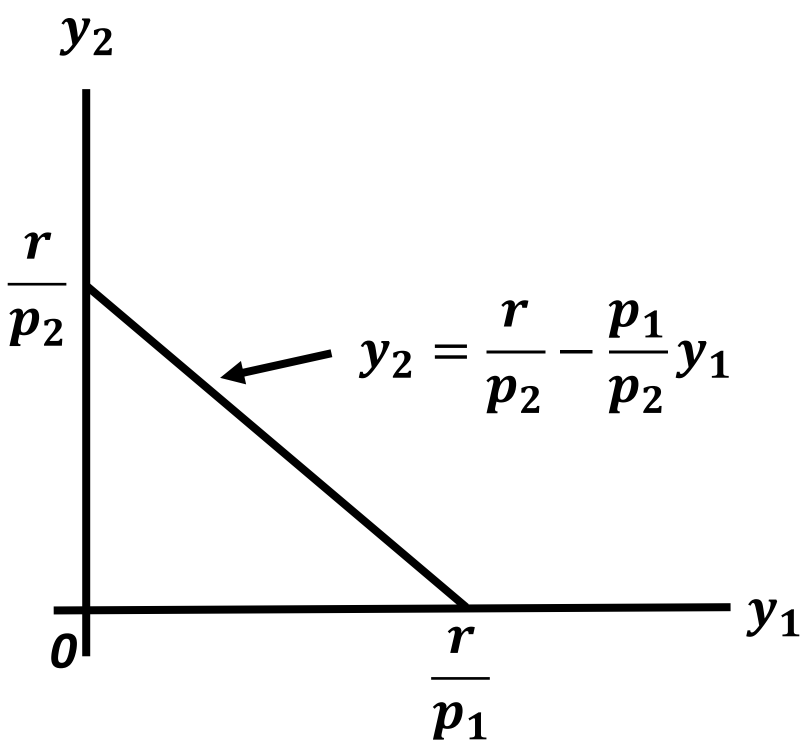
Definition 3. An *isorevenue line* (hyperplane) for (r, p), $\bar{\mathcal{R}}(r, p)$, is the set of output bundles that give a revenue of r at output prices p.

The mathematical definition of $\bar{\mathcal{R}}(r, p)$ is

$$\bar{\mathcal{R}}(r, p) \equiv \{y : r = py\}.$$

You should experiment with Figure 4 to determine how various permutations of changes in (r, p) change the visual representation of $\bar{\mathcal{R}}(r, p)$

Figure 4: Isorevenue Line $\bar{\mathcal{R}}(r, p)$



Profit maximizing (rational) producers

We now have the tools to examine the behavior of self-interested, price-taking producers. That behavior captures how producers respond to *market signals* in the form of prices (p, w) in the presence of feasibility constraints imposed by T. The producer's problem is to choose a technically feasible input-output bundle, $(x, y) \in T$, that makes profit, $py - wx$, as large as possible. In mathematical terms, producer's profit-maximization problem is written:

$$\pi(p, w) \equiv \max_{x,y} \{py - wx : (x, y) \in T\}.$$

Here $\pi(p, w)$ is the real-valued function of output and input prices that gives the maximum feasible profit for prices (p, w). As before, $\max_{x,y} \{\cdot\}$ is an instruction to search over x and y to find the largest element in the set $\{\cdot\}$ that contains profits, $py - wx$, conditional on $(x, y) \in T$. That means the producer seeks an $(x^*, y^*) \in T$ such that

$$py^* - wx^* \geq py - wx,$$

for all $(x, y) \in T$. We call $\pi(\cdot)$ the *profit function.*

We take a geometric approach and first treat the simple, single-input, single-output case. Figure 5 depicts the technology set, T, as an irregular heptagon. This shape may appear bizarre, but it was chosen purposely to ensure that T does not satisfy FDI, FDO, and convexity. That's done to emphasize that those properties are not essential to the argument.

Figure 6 augments Figure 5 by introducing the isoprofit line for (π^o, p^o, w^o), $\bar{\mathcal{P}}(\pi^o, p^o, w^o)$. The producer, as a price taker, cannot manipulate p^o or w^o. But, in the case illustrated, the producer can make a higher profit than π^o because a feasible input-output bundle, (x', y'), exists that lies on a higher isoprofit line. You can verify this by noting that (x', y') lies on the dotted line segment having the same slope as $\bar{\mathcal{P}}(\pi^o, p^o, w^o)$ that intersects the vertical axis above$\frac{\pi^o}{p^o}$ so that $p^o y' - w^o x' > \pi^o$. Therefore, π^o is not the largest profit available from T.

Instead, the producer maximizes profit for prices (p^o, w^o) at the point (x^*, y^*). The isoprofit line through (x^*, y^*), $\bar{\mathcal{P}}(p^o y^* - w^o x^*, p^o, w^o)$, shares a point in common with (intersects) T and defines the closed half space below it, $\mathcal{P}^-(p^o y^* - w^o x^*, p^o, w^o)$, that contains T

Figure 5: A Technology Set

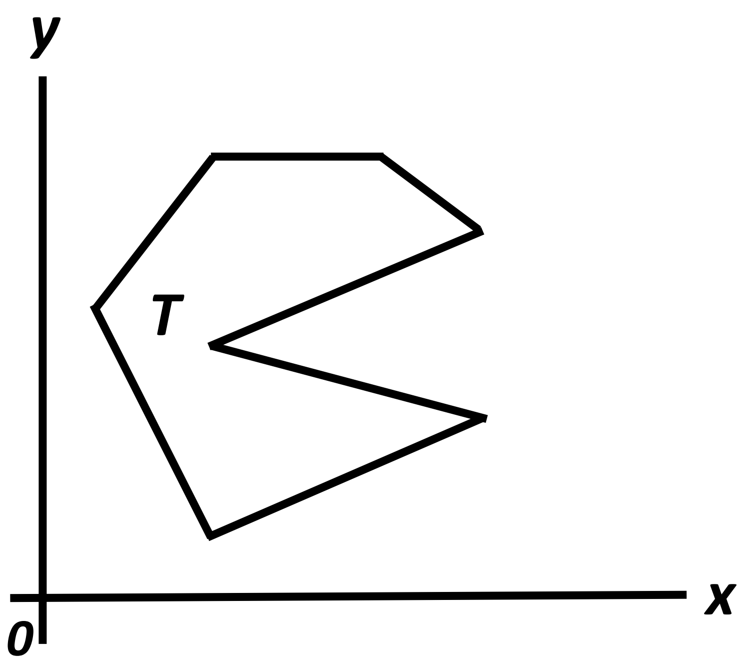
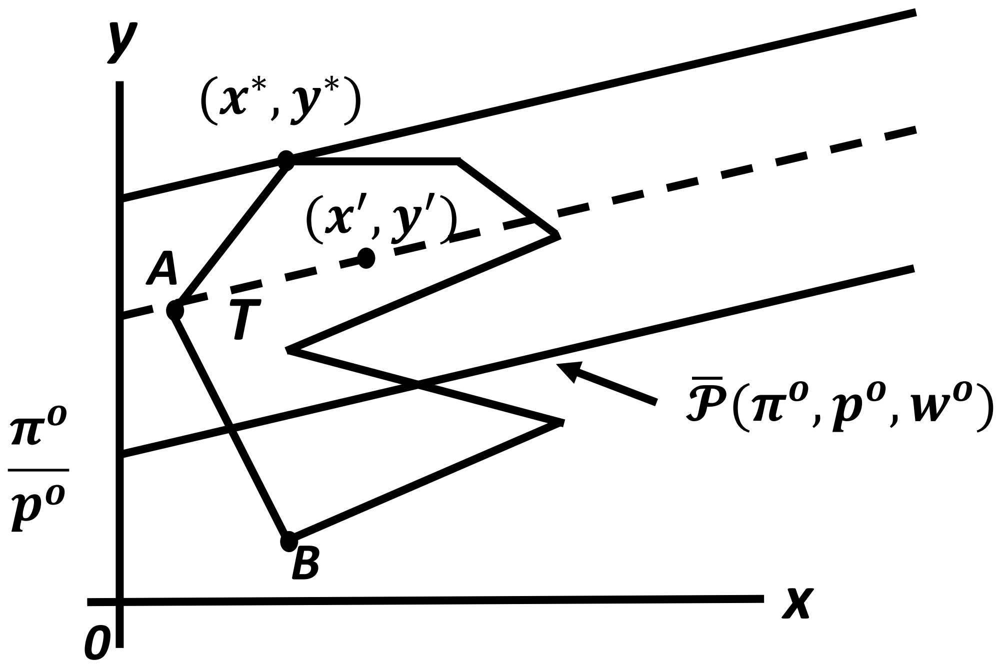


Figure 6: Maximizing Profit



as a subset. We write these conditions, respectively, in symbols as

$$(x^*, y^*) \in \bar{\mathcal{P}}\left(p^o y^* - w^o x^*, p^o, w^o\right) \cap T,$$

and

$$T \subset \mathcal{P}^-\left(p^o y^* - w^o x^*, p^o, w^o\right).$$

Together, they require that $(x^*, y^*) \in T$ and

$$p^o y^* - w^o x^* \geq p^o y - w^o x,$$

for all $(x, y) \in T$.

Let's see why the solution must occur at (x^*, y^*). Consider, for example, the consequences of producing at a point such as (x', y') that falls inside T. At that input-output bundle, the producer can keep the input fixed at x' and increase the output above y'. Because the input is held fixed, cost does not change. But increasing output increases revenue so that profit increases. That means (x', y') cannot be profit maximizing.

Another way to reach the same conclusion is to keep the output fixed at y', but to decrease x below x'. That leaves revenue unchanged but decreases cost. Again profit increases. Points such as (x', y') cannot be profit maximizing.

The general principle is:*Any input-output choice that falls on an isoprofit line that passes **inside** of the technology set cannot be profit maximizing.* Take a point such as A. Because it's on the boundary of T, it's technically feasible. But it also falls on an isoprofit line that passes inside the heptagon defining T. Because that isoprofit line passes inside the heptagon, it contains points such as (x', y') that allow profit-increasing moves. But because (x', y') and A lie on the same isoprofit line, they share a common profit, and A cannot be profit maximizing because feasible points exist that are more profitable than (x', y').

The producer maximizes profit by finding an input-output bundle where the isoprofit line through it touches T without going inside it. Because the isoprofit line just 'touches', or more euphemistically 'kisses', T's boundary at the solution, it is *tangent* to T's boundary. A tangency is needed. But a tangency is not enough. In addition, T can contain no points lying on a higher isoprofit line than the profit maximizer. T must, therefore, fall in the half space below the isoprofit line through (x^*, y^*).

Thus, the isoprofit line (hyperplane) for the solution, $\bar{\mathcal{P}}\left(p^o y^* - w^o x^*, p^o, w^o\right)$, bounds T from above while sharing a point with it. The hyperplane is said to *support T from above.* Perhaps, you've seen optimal solutions described as *tangencies* elsewhere in economics. That's not wrong, but more is required. You can experiment with Figure 6 to find other isoprofit lines with the same slope as the one through (x^*, y^*) that 'kiss' T, but involve a lower profit. (See Exercise 4.)

In characterizing the profit-maximizing solution we used $\in$ rather than $=$ in the statement

$$(x^*, y^*) \in \bar{\mathcal{P}}\left(p^o y^* - w^o x^*, p^o, w^o\right) \cap T.$$

That was intentional. Depending upon the shape of T and the structure of real prices, situations can arise where multiple elements of T solve the profit maximizing problem. Figure 7 illustrates a case where a continuum (an infinity) of points belonging to T are profit maximizing.

Profit Maximizers are Cost Minimizers and Revenue Maximizers

You can confirm in Figure 6 that $y^* = f(x^*)$ and $x^* = e(y^*)$. In the single-input, single-output setting that implies that y^* solves

$$R(p, x^*) \equiv \max_y \{py : (x^*, y) \in T\}$$

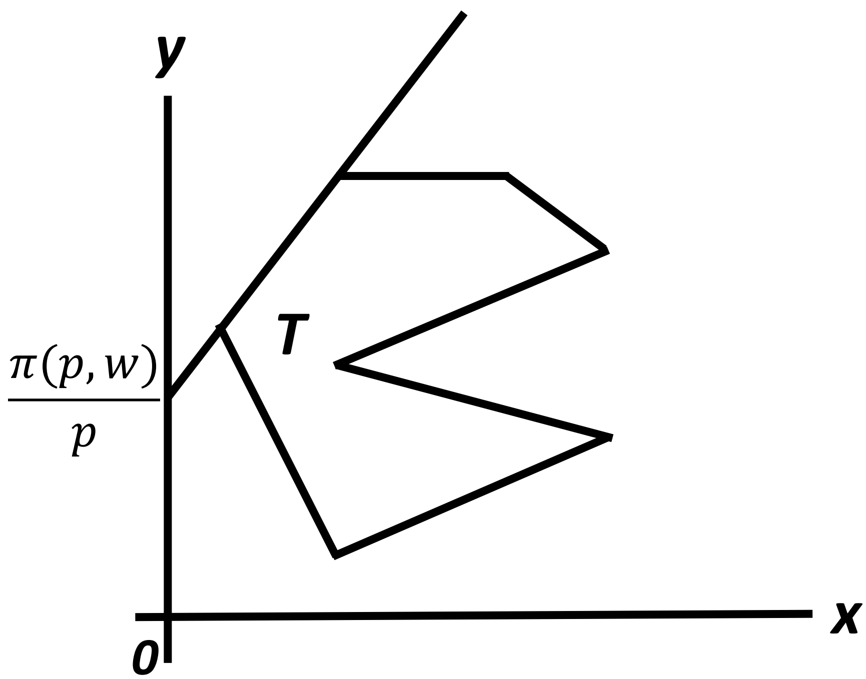
and that x^* solves

$$c(w, y^*) \equiv \min_x \{wx : (x, y^*) \in T\}.$$

That is, y^* maximizes the revenue that can be generated using x^*, and x^* is the cheapest feasible way to produce y^*. This result manifests the following general principle that applies regardless of the dimension of x and y.

Proposition 1. Profit maximizing producers always produce their optimal output at minimal cost and maximize the revenue received from their optimal input.

Figure 7: Multiple Profit-maximizing Solutions



To demonstrate Proposition 1, let (x^*, y^*) be a profit maximizing choice of inputs and outputs for (p, w) so that

$$\begin{aligned} \pi(p, w) &= \max_{x,y} \{py - wx : (x, y) \in T\} \\ &= py^* - wx^* \end{aligned}$$

The first statement in Proposition 1 claims that x^* is the cheapest way to produce y^*. Stated in mathematical terms, that requires

$$\min_x \{wx : (x, y^*) \in T\} = wx^*.$$

To verify, we use a *proof by contradiction*. That is, we assume that the statement we want to verify, in this case,

$$\min_x \{wx : (x, y^*) \in T\} = wx^*$$

is *false* and then show that assumption leads to a contradiction of known facts.

So, assume that the first part of Proposition 1 is false and that x^* is not the cheapest way to produce y^*. That means that another input bundle, call it x^o, must exist that can produce y^*, $(x^o, y^*) \in T$, and that is strictly cheaper than x^*,

$$wx^o < wx^*.$$

But this inequality implies that

$$py^* - wx^o > py^* - wx^* = \pi(p, w),$$

which says that $(x^o, y^*) \in T$ gives *more than the maximal feasible profit* $\pi(p, w)$. But this is impossible. We have arrived at a contradiction, and we conclude that profit maximizing x^* is cost minimizing for y^*.

A parallel argument, which we leave to you as Exercise 5, verifies the second part of Proposition 1. This proposition has important analytic consequences that require emphasis. It shows that we can recast the profit-maximization problem as:

$$\begin{aligned} \pi(p, w) &= \max_{x,y} \{py - wx : (x, y) \in T\} \\ &= \max_y \{py - c(w, y)\}, \end{aligned}$$

or as

$$\begin{aligned}\pi(p,w) &= \max_{x,y}\{py - wx : (x,y) \in T\} \\ &= \max_{x}\{R(p,x) - wx\}.\end{aligned}$$

Thus, three distinct, but equivalent, versions of the profit-maximization problem exist.

The first version is the one illustrated using Figure 6. The second is a two-stage approach. In the first stage (*cost minimization*), the producer chooses inputs to minimize the cost of producing a given output bundle, y. That is, the producer solves

$$c(w,y)) \equiv \min_{x}\{wx : (x,y) \in T\}$$

to obtain the *cost function*, $c(w,y)$, that gives the cheapest way to produce the output. Then in the second stage the producer solves the profit maximization problem by choosing the profit maximizing output bundle. The third version offers a different two-stage approach. In the third version, the first-stage *revenue maximization* problem is to choose outputs to maximize the revenue attainable from a fixed input bundle, x. So, the producer first solves

$$R(p,x) \equiv \max_{y}\{py : (x,y) \in T\}$$

to obtain the *revenue function*, $R(p,x)$, that gives the maximum possible revenue that can be generated from the input bundle x. Then in the second stage the producer chooses the profit maximizing input bundle.

These two-stage versions of the profit-maximization problem facilitate more focused economic analyses. The cost minimization case focuses on input interaction, and the revenue-maximization case focuses on output interaction. Both cases are examples of thought experiments used to simplify analysis. In the cost instance, output variation is “controlled” by holding y fixed permitting inputs to vary. In the revenue instance, the roles are switched. This analytic device mirrors a familiar strategy for solving complex decisions by breaking them into more easily digestible “sub-problems”. (These decompositions are special cases of a more general principle on maximization known as *Bellman’s Principle*, after the 20th century mathematician Richard E. Bellman.) We emphasize that the decomposition approach to solving the profit maximization is an *analytic strategy* that offers a tractable way for production economists to analyze producer behavior. The argument isn't that real-world producers are automatons who follow these two-stage algorithms, but that such analytic devices improve the economist's ability to understand and predict producer behavior.

Exercises

1. Show that if (π, p, w) vary proportionately so that $(\pi/p, w/p)$ remain unchanged, $\bar{\mathcal{P}}(\pi, p, w)$ remains unchanged (the isoprofit line is *homogeneous of degree zero* in (π, p, w)); if π increases holding (p, w) constant, the vertical intercept of the isoprofit line shifts up, and the slope remains unchanged resulting in a parallel upwards shift; if w increases while holding (π, p) constant, so that inputs become more expensive, the slope of the isoprofit line becomes steeper without changing the vertical intercept; and if p increases while holding (π, w) constant, the vertical intercept shifts downward and the slope becomes flatter.

2. Show that if (c, w_1, w_2) vary proportionately so that $(c/w_2, w_1/w_2)$ remain unchanged, $\bar{\mathcal{C}}(c, w)$ remains unchanged ($\bar{\mathcal{C}}(c, w)$ is homogeneous of degree zero in (c, w)); decreasing c while leaving (w_1, w_2) unchanged causes a *parallel shift* in the isocost line towards the origin; increasing w_1 holding c and w_2 constant causes the horizontal intercept of $\bar{\mathcal{C}}(c, w)$ to shift back towards the origin and the slope of the isocost line to become steeper (when viewed from the perspective of x_2); increasing w_2 while holding c and w_1 constant causes the horizontal intercept to shift towards the origin and the slope of the isocost cost line to become flatter.

3. Draw a picture that illustrates $\mathcal{C}^+(c, w)$ and $\mathcal{C}^-(c, w)$.

4. Explain why point B in Figure 6 is not profit maximizing.

5. Let (x^*, y^*) be profit maximizing. Prove that

$$\max_y \{py : (x^*, y) \in T\} = py^*,$$

Lectures on Neoclassical Production Economics[1]

Lecture 6: The Cost Function

Robert G Chambers [2]

January 6, 2024

[2]University of Maryland

Lecture Introduction

In this lecture we study producers who minimize the cost of producing their output. The core notion is the *cost function*, $c(w, y)$, introduced in Lecture 5. As Lecture 5 shows (see Figure 5.6), when the technology has a single input, $c(w, y) = we(y)$. So, one natural interpretation of $c(w, y)$ is as the *money-metric* generalization of the input-requirement function. The natural roadblock to generalizing $e(y)$ to multiple inputs is our inability to add different inputs in a meaningful manner. The cost function removes that obstacle by converting inputs to monetary values and then adding.

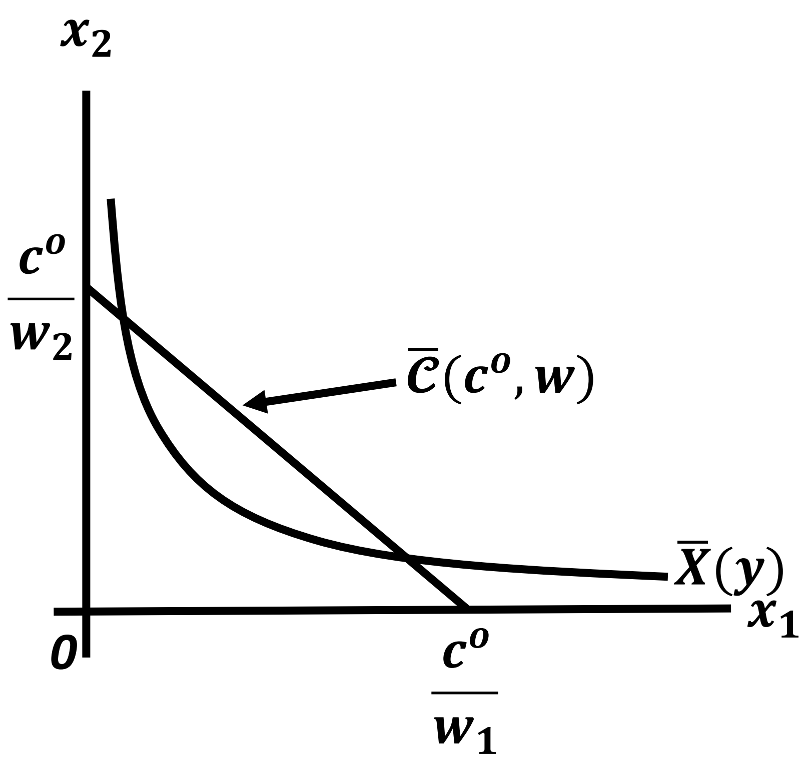
We first illustrate the solution to the cost-minimization problem using the style of argument introduced in Lecture 5. Then we characterize $c(w, y)$'s *behavior as a function of the input prices, w.* We emphasize that this characterization *only* requires one restriction on T, which is that $X(y) \neq \emptyset$. We close the lecture by examining the consequences for $c(w, y)$, as a function of outputs, y, of imposing restrictions upon T.

A Visual Perspective on Cost Minimization

Figure 1 depicts two objects in the (x_1, x_2) plane. The first is a nonempty input set with isoquant, $\bar{X}(y)$. It's drawn to be smooth, consistent with convexity, and to satisfy FDI. (These are not essential to the argument.) The second object is the isocost line for a cost of c^o dollars and input prices w, $\bar{\mathcal{C}}(c^o, w)$. The isocost line and the input set intersect, $\bar{\mathcal{C}}(c^o, w) \cap X(y) \neq \emptyset$. So, points on the isocost line fall in the input set, and the producer can spend c^o dollars and afford to produce y.

There are input bundles that fall below $\bar{\mathcal{C}}(c^o, w)$ but above $\bar{X}(y)$. Such input bundles can produce y but fall inside $\mathcal{C}^-(c^o, w)$ and so cost less than c^o. c^o is not the minimal cost to produce y. Following the arguments used in Lecture 5, a cost-minimizing solution cannot occur inside of $X(y)$ or at a point on $\bar{X}(y)$ that is on an isocost line that passes inside of $X(y)$. The cost minimizing solution, call it x^*, occurs where its isocost line, $\bar{\mathcal{C}}(wx^*, w)$, touches $\bar{X}(y)$ and $X(y)$ is contained in the half space above it, $\mathcal{C}^+(wx^*, w)$. That is, x^* is

Figure 1: The Minimal Cost Problem

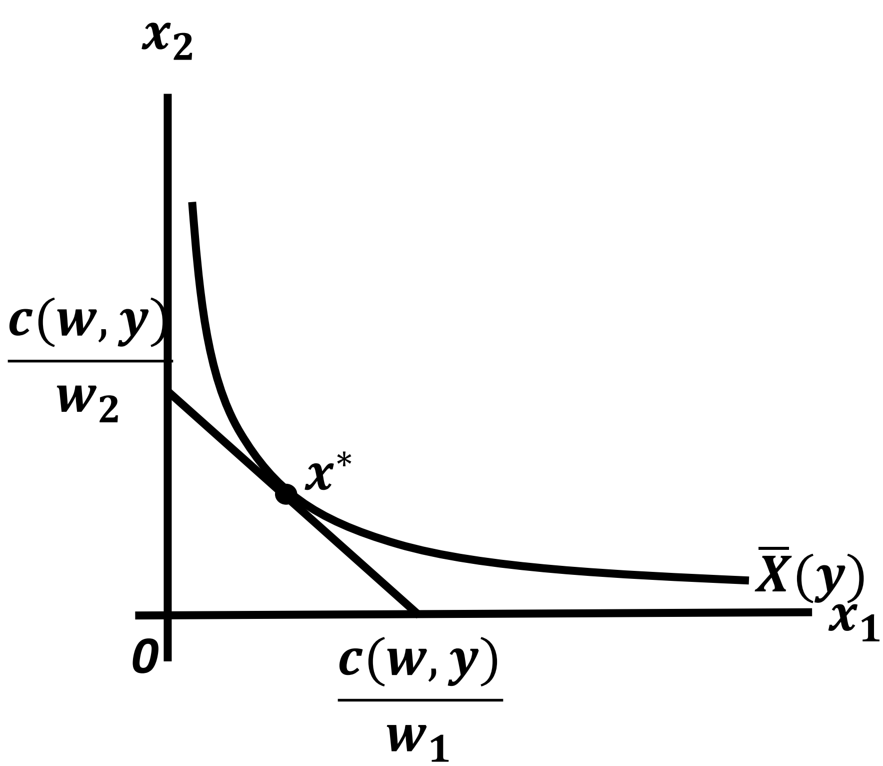


cost minimizing for y if and only if

$$
\begin{aligned}
x^* &\in X(y) \cap \bar{\mathcal{C}}(wx^*, w) \text{ and} \\
X(y) &\subset \mathcal{C}^+(wx^*, w)
\end{aligned}
$$

so that $\bar{\mathcal{C}}(wx^*, w)$ *supports* $X(y)$ *from below.*[1] The input set is tangent to and 'sits' on the isocost line. The implied tangency between $\bar{X}(y)$ and $\bar{\mathcal{C}}(c, w)$ gives the rule-of-thumb for cost minimization: *the marginal rate of substitution equals (minus) the input-price ratio.* Figure 2 illustrates.

Figure 2: A Minimal-Cost Solution



[1]Recall from Chapter 5 that the requirement for profit maximization is that the isoprofit line support T from above. We seek to minimize cost so the logic and intuition is reversed and support must be from below.

The Cost Function and Its Properties in Input Prices

Definition 1. The *cost function*, $c(w, y)$, for input prices, w, and output bundle, y, gives the minimal cost of producing y.

The mathematical version of this definition reiterates the one offered in Lecture 5:[2]

$$\begin{aligned} c(w, y) &\equiv \min_x \{wx : (x, y) \in T\} \\ &= \min_x \{wx : x \in X(y)\}. \end{aligned}$$

Provided that y can be produced, that is, $X(y) \neq \emptyset$, $c(w, y)$ is well defined. We state its properties as a function of input prices, w. Then we explain and discuss them.

Cost Function Properties in Input Prices

- $c(w, y) \geq 0$ ($c(w, y)$ nonnegative)
- $w' \geq w \Rightarrow c(w', y) \geq c(w, y)$ (nondecreasing in w)
- $c(\mu w, y) = \mu c(w, y) \quad \mu > 0$ (positively homogeneous in w)
- $c(\lambda w^o + (1 - \lambda) w', y) \geq \lambda c(w^o, y) + (1 - \lambda) c(w', y) \quad \lambda \in (0, 1)$ (concave in w)

$c(w, y)$ **nonnegative** This property follows from measuring inputs and input prices as nonnegative real numbers. Cost, as the sum of input prices times input use, must be nonnegative.

$c(w, y)$ **is nondecreasing in w** Let x' be the cheapest way to produce y for w', so that

$$w'x' = c(w', y).$$

Now let input prices fall to w from w' so that $w' \geq w$. Because $X(y)$ does not depend upon input prices, x' can still produce y. Thus, "doing nothing" is a feasible response to the input price change. The cost of the "do-nothing" response is

$$wx' \leq w'x'.$$

[2] The cost function, $c(w, y)$, in math terms is the *support function* for the set $X(y)$.

The best producer response to input prices falling cannot cost more than the "do-nothing" response so that

$$c(w, y) = \min\{wx : x \in X(y)\} \leq wx'.$$

Combining these arguments gives

$$c(w', y) \geq wx' \geq c(w, y)$$

for $w' \geq w$, as required.

If inputs become more expensive, the minimal cost of producing an output cannot decrease. When depicted in (w, c) space, the graph of $c(w, y)$ can have a positive or zero slope, but never negative.

Positive homogeneity in w (only real prices matter) The aphorism 'only real prices matter' summarizes this property. *Producers do not respond to nominal input price changes that leave real input prices unchanged.*

Look at Figure 1 and consider multiplying all input prices, w, by a common positive number, $\mu > 0$. That causes the vertical and horizontal intercepts of the isocost line to shift towards the origin if $\mu > 1$ and towards ∞ if $\mu < 1$. But its slope is unchanged because

$$-\frac{\mu w_1}{\mu w_2} = -\frac{w_1}{w_2}.$$

As Figure 2 illustrates, the key to where the producer locates on $\bar{X}(y)$ is where its slope equals the isocost's slope $-\frac{w_1}{w_2}$. Call the solution x^*. Because $-\frac{w_1}{w_2}$ is constant for nominal price changes, x^* stays a solution when all input prices are multiplied by a common multiple. Nominal costs change to $\mu wx^* = \mu c(w, y)$, however, as nominal input prices move to μw from w.

The formal version of this argument invokes the definition of $c(w, y)$ at $\mu > 0$:

$$\begin{aligned} c(\mu w, y) &= \min\{\mu wx : x \in X(y)\} \\ &= \mu \min\{wx : x \in X(y)\} \\ &= \mu c(w, y) \end{aligned}$$

The second equality is the key. It recognizes that making μwx as small as possible by varying x, with μ a fixed positive number, requires making wx as small as possible. And so, if x^* makes wx as small as possible, it does the same for μwx.

One way to think about why 'only real prices matter' is to think of re-stating prices denominated in US dollars ($) in terms of cents. Moving from dollars to cents multiplies nominal prices by 100, a special case of $\mu > 0$. But in terms of real buying power, nothing changes.

$c(w, y)$ **is concave in** w Before showing that $c(w, y)$ is concave in w, let's see its economic meaning. We know that the graph of $c(w, y)$ in (w, c) space does not have a negative slope. So, our intuition is that $c(w, y)$ rises as input prices rise. From our discussion of concave production functions in Lecture 3, we know that concavity in w means that $c(w, y)$'s *slope* is nonincreasing in w. Figure 3 illustrates the case where it is decreasing.

Pick an arbitrary input price, w^*, and mark off its cost on the vertical axis in Figure 3,

$$c(w^*, y) = w^* x^*,$$

where x^* now represents the cost minimizing input bundle when input prices are w^*. When input prices change, the producer must decide how to adjust the input bundle. Because input price changes do not affect $X(y)$, the only technical requirement is that the adjustment keeps the input bundle in $X(y)$. "Doing nothing" is a feasible response. So, let prices change to w^o and assume that the producer does nothing. The producer's cost is now $w^o x^*$, and the cost change is $(w^o - w^*) x^*$. We illustrate the "do-nothing" response in Figure 4 by moving along the line segment connecting $(w^*, w^* x^*)$ and $(w^o, w^o x^*)$, whose slope is x^*. (Exercise 2 asks you to to verify that similar reasoning applies for price rises.)

Because "doing nothing" is a feasible reaction, it places an *upper bound* on the producer's best response to the price change. Therefore, the minimal cost at the new prices w^o cannot be larger than $w^o x^*$, that is

$$c(w^o, y) \leq w^o x^*.$$

But that's what a cost function concave in w predicts as Figure 4 shows.

Why is this true? In Figure 4, we assume that the producer does not respond to the input-price change. This is possible, but is it expected? Let's return to Figure 2 and consider how

Figure 3: Cost Concave in w

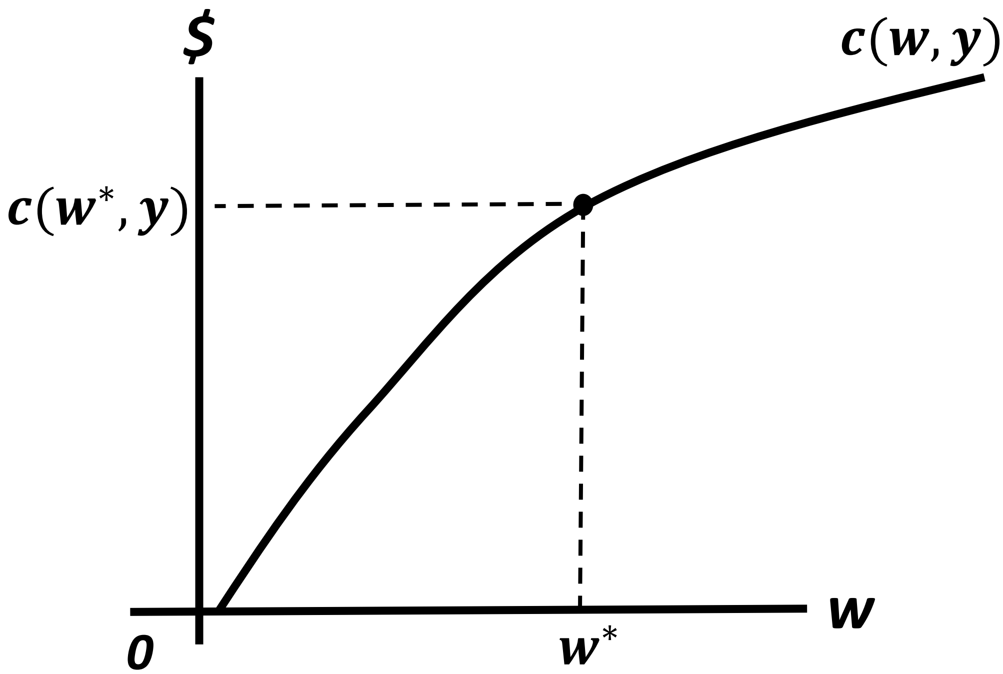
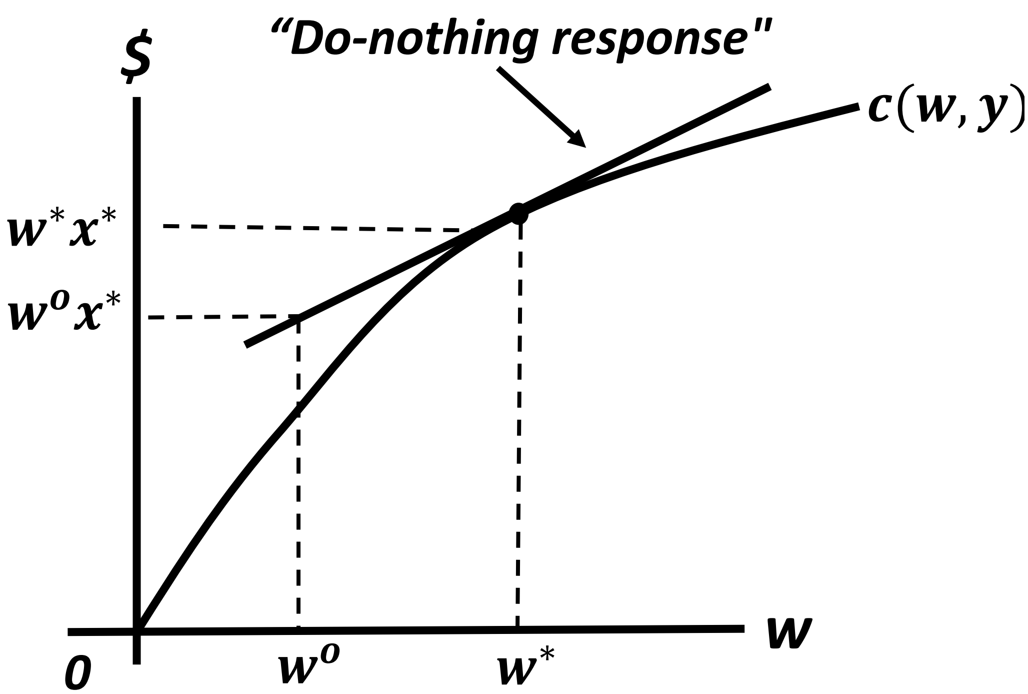
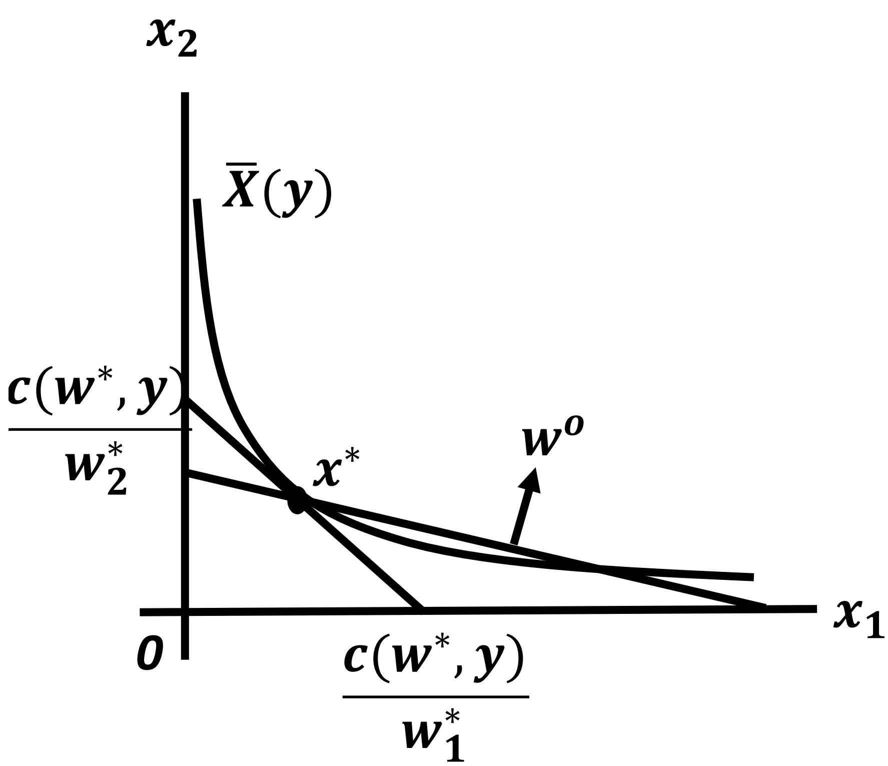


Figure 4: Shephard's Lemma



that solution changes if prices change. If both prices change proportionately, no adjustment is made. But if the price of input 1 falls by 50 per cent, the slope of the isocost line becomes flatter (from the perspective of the x_2 axis). The isocost line through the "do-nothing" option now passes inside $X(y)$. As Figure 5 illustrates, cheaper alternatives are available. (In Figure 5, the line segment labelled with a perpendicular arrow w^o is the isocost line for the prices w^o.) Thus, if the producer can substitute more of the now cheaper input 1 for

Figure 5: Price Change Causes Substitution



input 2, costs can be decreased beyond the "do-nothing" option.

Concavity of the cost function in input prices manifests the producer's ability to substitute between inputs in response to price changes.

Verifying these arguments in arbitrary dimensions is easy, but requires some algebra (multiplication and addition) and reasoning. We want to compare three values, $c(\lambda w^o + (1-\lambda) w', y)$, $c(w^o, y)$, and $c(w', y)$. Let x^* represent the input bundle that solves the cost-minimizing problem for a weighted average of w^o and w', $\lambda w^o + (1-\lambda) w'$ for some $\lambda \in (0,1)$. That is,

$$(\lambda w^o + (1-\lambda) w') x^* = c(\lambda w^o + (1-\lambda) w', y).$$

Because x^* is the solution to the cost-minimizing problem for this weighted average price, $\lambda w^o + (1-\lambda) w'$, it must belong to $X(y)$. But that means

$$\begin{aligned} w^o x^* &\geq c(w^o, y) \quad \text{and} \\ w' x^* &\geq c(w', y). \end{aligned}$$

x^* must be at least as expensive at input prices w^o and w^* as the respective cost minimizing bundles. Multiplying both sides of the first inequality by $\lambda > 0$, both sides of the second by $1 - \lambda > 0$, and adding the results gives

$$(\lambda w^o + (1-\lambda) w') x^* \geq \lambda c(w^o, y) + (1-\lambda) c(w', y).$$

The left hand side of this inequality equals $c(\lambda w^o + (1-\lambda) w', y)$. Making that substitution, we have

$$c(\lambda w^o + (1-\lambda) w', y) \geq \lambda c(w^o, y) + (1-\lambda) c(w', y)$$

for $\lambda \in (0,1)$, which is what we wanted to show. This concludes the argument.

Cost Minimizing Input Demands

Just as we are interested in how costs behave as input prices change, so too are we interested in how *cost minimizing demands* respond to price changes.

Definition 2. A *cost minimizing demand vector*, $x(w, y)$, for input prices w and output y is any solution to the cost minimization problem for w and y.

The mathematical version of the definition is

$$x(w, y) \in \underset{x}{\operatorname{argmin}} \{wx : x \in X(y)\}.$$

Here the expression "$\text{argmin}_x \{\cdot\}$" can be written in words as "any x that gives the minimal value in the set $\{\cdot\}$". We intentionally use $\in$ rather than $=$ because there can be multiple input bundles that provide the same minimal cost (see Exercise 3).

Our focus is on $x(w,y)$ as a function of w *holding* y *constant*. We start by repeating a basic observation. Because $x(w,y)$ is cost minimizing for y, it must belong to $X(y)$. Therefore, if we choose an arbitrary w^*

$$w^o x(w^*,y) \geq c(w^o,y)$$

for all possible w^o, and

$$w^* x(w^*,y) = c(w^*,y).$$

In words, $w^o x(w^*,y)$ is always greater than or equal to the minimal cost for input prices w^o, $c(w^o,y)$, and equals it when $w^o = w^*$. Subtracting the second expression from the first gives

$$(w^o - w^*) x(w^*,y) \geq c(w^o,y) - c(w^*,y) \tag{1}$$

for all w^o as a condition that cost-minimizing demands must satisfy.[3] The intuition is that the cost change associated with doing nothing as input prices change is always larger than the change in minimal cost.

The expression,

$$w^o x(w^*,y) \equiv \sum_{n=1}^{N} w_n^o x_n(w^*,y),$$

where $x_n(w^*,y)$ represents the cost-minimizing demand for the *nth* input, is a *linear function* of the input prices w^o whose slopes equal the respective cost minimizing demands, $(x_1(w^*,y), x_2(w^*,y), \ldots, x_N(w^*,y))$. We have just shown that the cost function, $c(w^o,y)$, is a concave function of the input prices w^o. So, we can rephrase

$$w^o x(w^*,y) \geq c(w^o,y),$$

and

$$w^* x(w^*,y) = c(w^*,y)$$

[3]In mathematical jargon, expression (1) says that the input demand, $x(w^*,y)$, is a *supergradient for* $c(w,y)$ *at* w^*.

for all w^o as requiring that the linear function $w^o x(w^*, y)$ provides an upper bound to the concave function $c(w^o, y)$ that equals it when $w^o = w^*$. In geometric terms, the graph of the linear function, $w^o x(w^*, y)$, always lies above the graph of the concave function, $c(w^o, y)$, in (w^o, c) space and coincides with it at $(w^*, c(w^*, y))$.

To visualize, consider Figure 4 and let $x(w^*, y) \equiv x^*$ in the figure. The graph of the linear function, $w^o x^*$ in the (w^o, c) plane is illustrated by the "Do-nothing" line segment passing through $(w^*, c(w^*, y))$ and $(w^o, w^o x^*)$. The "Do-nothing" line segment lies everywhere above the graph of the cost function and just touches (is tangent to) it at $(w^*, c(w^*, y))$. We conclude that the cost-minimizing demand for the input at input prices w^* is given by the slope of the "Do-nothing" line tangent to the graph of $c(w, y)$ at $(w^*, c(w^*, y))$.

Our argument works regardless of the number of inputs. When a linear function has more than one argument, its graph is not referred to as a line but as a *hyperplane* (plane works for two dimensions, the hyper is for three dimensions and above), and the N-dimensional analogue of the slope is a vector referred to as its *normal*. Thus, for the linear function of w^o,

$$w^o x(w^*, y) \equiv \sum_{n=1}^{N} w_n^0 x_n(w^*, y),$$

its *normal* is the vector

$$(x_1(w^*, y), x_2(w^*, y), ..., x_N(w^*, y)) \equiv x(w^*, y)$$

of cost minimizing demands at w^*.

These arguments yield a famous result in economics:[4]

Shephard's Lemma For all w^* and y, any solution to the cost-minimization problem, $x(w^*, y)$ must be the normal for a hyperplane tangent to the graph of $c(w, y)$ at $(w^*, c(w^*, y))$.

[4]You may encounter another version of this result elsewhere, which runs: The vector of cost minimizing demands at w^* equals the gradient of the cost function at w^*. That isn't wrong, but it requires additional structure upon T. Thus, it is less general. The more precise version that applies generally is that the vector of cost minimizing demands is a *supergradient* of the cost function at w^*, which is what our statement requires. See, also, footnote 3.

Overstating the importance of Shephard's Lemma is hard. It offers a way to recapture cost-minimizing demands. And *because those input demands can be captured from* $c(w, y)$*, they inherit their properties in* w *from* $c(w, y)$ as we now show.

Positive homogeneity of the cost function in input prices, $c(\mu w, y) = \mu c(w, y)$, implies that cost-minimizing demands don't respond to nominal price changes that leave real prices unchanged. Let's look at this using Shephard's Lemma. For input prices w^* the cost minimizing input demand vector, $x(w^*, y)$, satisfies expression (1). Now suppose that we hold $x(w^*, y)$ constant but let input prices w^* and w^o in (1) move to μw^* and μw^o. Will $x(w^*, y)$ still satisfy (1)? The answer is yes because positive homogeneity of cost ensures that

$$c(\mu w^o, y) - c(\mu w^*, y) = \mu (c(w^o, y) - c(w^*, y)).$$

Thus,

$$\begin{aligned}
(w^o - w^*) x(w^*, y) &\geq c(w^o, y) - c(w^*, y) \\
&\Updownarrow \\
\mu (w^o - w^*) x(w^*, y) &\geq \mu [c(w^o, y) - c(w^*, y)] \\
&\Updownarrow \\
(\mu w^o - \mu w^*) x(w^*, y) &\geq c(\mu w^o, y) - c(\mu w^*, y)
\end{aligned}$$

for $\mu > 0$. Therefore, $x(w^*, y)$ remains cost-minimizing for μw^*, whence

$$x(\mu w, y) = \mu c(w, y) \quad \mu > 0.$$

Cost minimizing input demands *are homogeneous of degree zero in input prices,* w*,* which is the mathematical version of 'only real prices matter'.

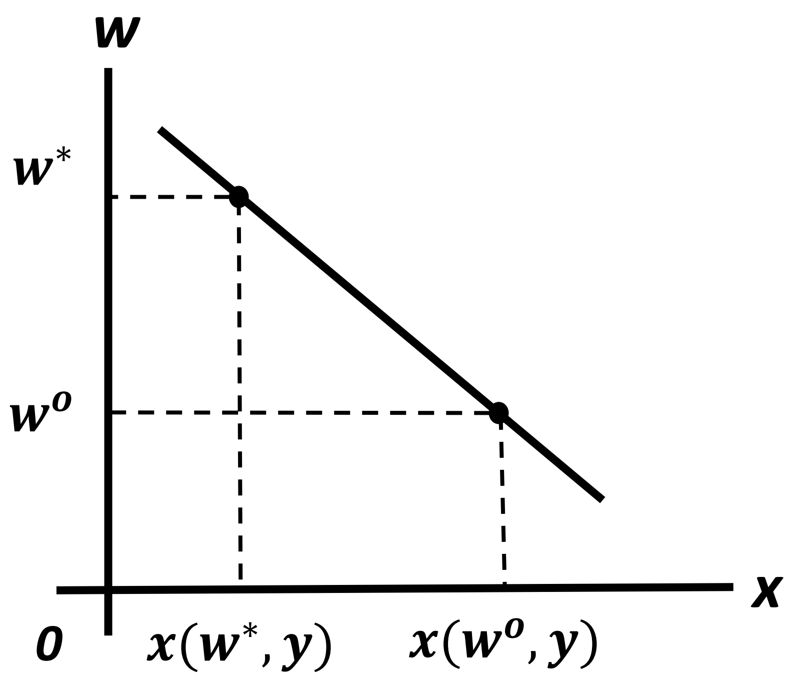
Figure 4 illustrates the consequences of *real* price changes for cost minimizing input demands. The graph of $c(w, y)$ is steeper at w^o than at w^*. Applying Shephard's Lemma gives $x(w^o, y) \geq x(w^*, y)$ or

$$(x(w^o, y) - x(w^*, y))(w^o - w^*) \leq 0. \tag{2}$$

The input price and the cost minimizing demand vary inversely. Graphing $x(w, y)$ in the (x, w) plane gives a cost-minimizing *input demand curve* with a negative slope (see Figure 6).[5]

[5]Here we note a technical glitch. The mathematical convention is that the "independent" variable is

Figure 6: Cost Minimizing Input Demand Curve



We confirm the visual story with a simple but rigorous argument. (It holds regardless of the dimension of x and w.) Take two cost minimizing demand bundles, $x(w^*, y)$ and $x(w^o, y)$, for the same output y but for different price vectors, w^* and w^o. Because they are cost minimizing demands for y, both must belong to $X(y)$. And although each is cost minimizing for its respective input prices, they are not cost minimizing for the others. That is,

$$\begin{aligned} w^* x(w^o, y) &\geq w^* x(w^*, y) \\ w^o x(w^*, y) &\geq w^o x(w^o, y) \end{aligned}$$

Adding these inequalities together gives

$$w^* x(w^o, y) + w^o x(w^*, y) \geq w^o x(w^o, y) + w^* x(w^*, y),$$

and collecting terms gives

$$(x(w^o, y) - x(w^*, y))(w^o - w^*) \leq 0,$$

which repeats (2) that we obtained from Figure 4.[6]

represented along the horizontal axis and the "dependent" variable on the vertical axis. Because we represent demand for x as a function of w, w is the independent variable and x the dependent. Nevertheless, the tradition in economics is to plot demand curves with price on the vertical axis and quantity on the horizontal axis. To avoid confusion, we adhere to the economics convention.

[6]Expression (2) is an important special case of a more general result. Consider, three input-price vectors, w^1, w^2, and w^3 and their respective input demands. By the definition of a cost-minimizing demand, it must be true that

$$\begin{aligned} w^1 x(w^1, y) &\leq w^1 x(w^2, y) \\ w^2 x(w^2, y) &\leq w^2 x(w^3, y) \\ w^3 x(w^3, y) &\leq w^3 x(w^1, y). \end{aligned}$$

Adding gives

$$w^1(x(w^1, y) - x(w^2, y)) + w^2(x(w^2, y) - x(w^3, y)) + w^3(x(w^3, y) - x(w^1, y)) \leq 0.$$

And repeating the same argument for w^1, w^2, w^3,...,w^K, with K arbitrary gives

$$w^1(x(w^1, y) - x(w^2, y)) + w^2(x(w^2, y) - x(w^3, y)) + ... + w^K(x(w^K, y) - x(w^1, y)) \leq 0.$$

Writing (2) in expanded form gives

$$\sum_{n=1}^{N} \left(x_n\left(w^o, y\right) - x_n\left(w^*, y\right)\right)\left(w_n^o - w_n^*\right) \leq 0. \tag{3}$$

Cost-minimizing input use tends to vary inversely with input prices. If only the price for input k changes so that $(w_n^o - w_n^*) = 0$ for $n \neq k$ expression (3) becomes

$$\left(x_k\left(w^o, y\right) - x_k\left(w^*, y\right)\right)\left(w_k^o - w_k^*\right) \leq 0.$$

The kth cost minimizing demand is nonincreasing in the kth input price. The choice of k was arbitrary, so the result must hold for all cost minimizing demands. Cost minimizing demands must be nonincreasing functions of their *own price*, where the *own price* for input k is w_k. As elsewhere, we use the term nonincreasing to be precise. But the intuition is that cost minimizing demands are decreasing in their own prices and cannot slope upward. There are no "Giffen" input demands.

How do we expect cost-minimizing input demands to respond to changes in the prices of *other* inputs. Answering this question carries us beyond what our theory predicts. We can say some things about what happens, but questions remain.

For example, we observed in Figure 5 that halving the price of input 1 led to an increase in the use of input 1 and a decrease in the use of input 2. The demand for input 2 fell as the price of input 1 fell. We referred to that as input 1 substituting for input 2. When there are only two inputs (which is unrealistic in most instances), this pattern must apply under FDI. As the real price between them changes, one input's use must decrease while the others must increase. So, if we look at Figure 6, and assume that the input and input price depicted in the graph are x_2 and w_2, halving w_1 leads to the demand curve for x_2 to shift inwards. Less of x_2 is demanded at every w_2 when the two inputs are substitutes after w_1 falls.

From the perspective of input 2, economists call the price of input 1 a *cross price* and the induced adjustment of input 2 a *cross-price demand adjustment.* In general, for input k, a change in any input price, w_j, with $j \neq k$ is a *cross-price change.* When inputs k and j

as a consequence of cost minimization. This property of cost minimizing demands is called *cyclical monotonicity in* w, and it along with homogeneity of degree zero exhaustively characterizes the behavior of cost minimizing demands in w.

have *positive* cross-price adjustments, they are called *input substitutes.* When inputs k and j have *negative* cross-price adjustments, they are *input complements.*

The intuitive story is that input demands vary inversely with their own prices. Thus, if its own price falls, more of the input is used. Under FDI, that suggests to keep output, y, constant some other input's use must fall as the producer moves along the isoquant, $\bar{X}(y)$. Those that do are the *substitutes.* But FDI does not require that all other inputs decrease.

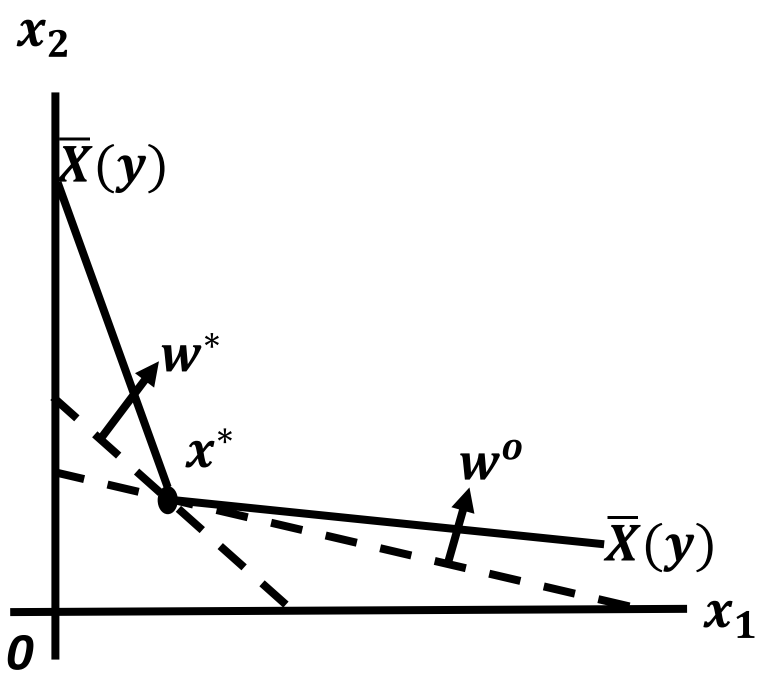
A concrete example illustrates. Think of a farmer who grows and harvests corn using different kinds of capital equipment (tractors, sprayers, combines), labor, land, fertilizers, pesticides, and other inputs. Now suppose that the price of land falls. How would you expect the farmer to adjust input use to keep output constant? Without empirical work, we can't know for sure, but it's easy to imagine instances where the farmer would use more labor. Land and labor would then be complements.

We close this discussion with a brief look at a canonical technology for which an input price change causes no change in input use (also see Exercise 3). In Figure 7, the piecewise linear curve $\bar{X}(y)\, x^* \bar{X}(y)$ depicts an isoquant. If the input set for y, $X(y)$, is everything on or above this piecewise linear object, it satisfies nonemptiness, FDI, and convexity. But we have drawn in two separate isocost lines, one, w^*, with a higher real price of input 1, $\frac{w_1}{w_2}$, than the other w^o that generate the same x^* as the cost minimizing choice.

Is Figure 7 realistic? Judging from observed behavior, *maybe* is a good answer. If you compare Figures 5 and 7, you will note that Figure 5 suggests that producers respond to even tiny price changes by altering their input usage. That often doesn't happen. Producers often make no input adjustments even for relatively large price changes. Figure 7 explains such behavior. Some who have been seduced by drawing smooth isoquants might believe such behavior contradicts our theory.[7] Figure 7 shows that's wrong.

[7]This is one area where economists' predilection for calculus-based arguments has caused needless confusion.

Figure 7: Input Price Change Causes No Demand Change



Properties of the cost function in y

The properties of the cost function in input prices, including Shephard's Lemma, follow from the type of decisionmakers we study and the environment in which they operate and not from specific restrictions upon T (for example, FDI or convexity). Assuming that producers are price takers means that they treat costs incurred by purchasing inputs as a *linear function* of inputs. That, in turn, means that the set of input bundles costing the same amount, what we call $\bar{\mathcal{C}}(c, w)$, is a line (hyperplane). The profit-maximization assumption implies that producers minimize cost and seek to operate on the lowest isocost line (hyperplane) consistent with producing y. Their optimal choice must be separated from other feasible alternatives by a linear inequality from which the properties of $c(w, y)$ flow.

The assumptions about the setting, however, do not determine how minimal costs vary with output. To make those kind of predictions, we need more. We now examine how imposing assumptions on $X(y)$ affects the cost function's behavior as a function of y.

Cost Function Properties in Outputs

- $FDO \Rightarrow c(w, y') \geq c(w, y), \;\; y' \geq y$ (nondecreasing in y)
- Convexity of $T \Rightarrow c(w, \lambda y^o + (1-\lambda) y') \leq \lambda c(w, y^o) + (1-\lambda) c(w, y') \;\; for \; \lambda \in (0, 1)$ (convex in output)
- $NFL \Rightarrow c(w, y) > 0 \; for \; y \geq 0, y \neq 0$. $NFC \Rightarrow c(w, 0) = 0$.

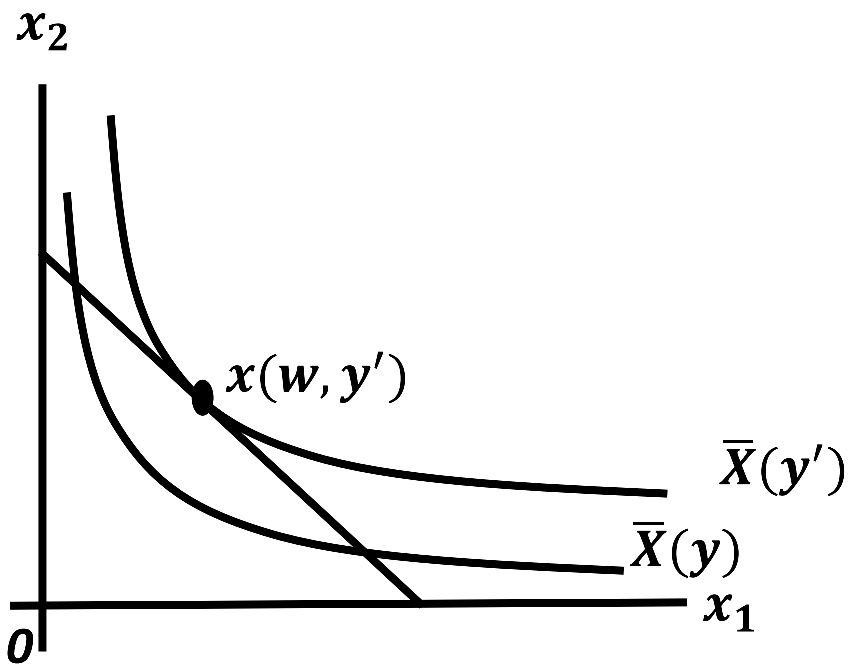
FDO implies costs are nondecreasing in output Recall that FDO requires that input sets get smaller as outputs expand, or said another way, they get larger as outputs decrease. We wrote this as

$$FDO \Rightarrow X(y') \subset X(y)$$

for $y' \geq y$. Assume, therefore, that we have identified a cost-minimizing input bundle for y', $x(w, y')$. By FDO, $x(w, y') \in X(y)$ for $y' \geq y$, the cost minimizing input demand for y' ("doing nothing") is a feasible response to output moving from y' to y. Doing nothing keeps cost constant. Therefore, a cost-minimizing producer, having this option, cannot respond by increasing cost.

Figure 8 illustrates the argument. There $\bar{X}(y)$ and $\bar{X}(y')$ depict the isoquants for y and

Figure 8: FDO and Minimal Cost



y'. The cost minimizing solution for y' is at $x(w, y')$. The isoquants have been drawn so that $X(y')$ is contained in $X(y)$, $x(w, y')$ can produce y but is not cost minimizing. Therefore, *marginal cost* is nonnegative, where *marginal cost* is the cost-analogue of *marginal product.* In the single-output case, it's the change in cost divided by a change in output. So for 'tiny' output changes, it's $c(w, y)$'s slope in the (y, c) plane. In the multiple-output case, the marginal costs of the different outputs at, for example, y^* are given by the elements of the normal to the hyperplane tangent to the graph of $c(w, y)$ at $(y^*, c(w, y^*))$.

Convexity implies the cost function is convex in output You will recall from Lecture 3 that a convex T ensures that the input-requirement function, $e(y)$, is convex as a function of y. (See Figure 2, Lecture 3.) The cost function is the money-metric generalization of the input-requirement function, it, too, is convex as a function of y for convex T.

Marginal cost curves graphed in $(y, \$)$ space for a cost function convex in y have a positive slope that rises as output rises. (Exercise 5 asks you to explain why this is true and to draw the appropriate pictures.) This is viewed as the normal state of affairs and many (but not all) visual depictions of marginal cost are drawn this way.[8]

We use a familiar strategy to show this result. Take minimizing input bundles for two arbitrary outputs bundles, y^o and y', and denote them as $x(w, y^o)$ and $x(w, y')$. T a convex set requires that any weighted average of these bundles,

$$\lambda x(w, y^o) + (1-\lambda) x(w, y'),$$

can produce the corresponding weighted average of y^o and y', $\lambda y^o + (1-\lambda) y'$. But that means its associated cost is at least as large as minimal cost,

$$c(w, \lambda y^o + (1-\lambda) y') \leq w(\lambda x(w, y^o) + (1-\lambda) x(w, y')).$$

Bringing w inside the parentheses on the right-hand side of this expression and using the definition of cost-minimizing demands gives

$$c(w, \lambda y^o + (1-\lambda) y') \leq \lambda c(w, y^o) + (1-\lambda) c(w, y') \;\; for \; \lambda \in (0,1),$$

as desired.

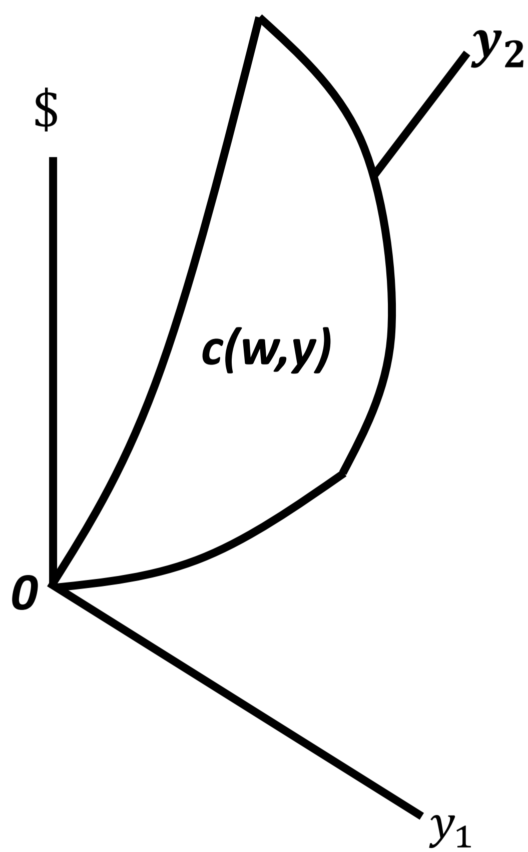
No Free Lunch and No Fixed Costs The names of these properties evoke their consequences for $c(w, y)$. NFL means what it says, you can't get something for nothing. To produce a positive output, you must incur a positive cost. It is a direct consequence of requiring a positive input to produce a positive output. NFC, on the other hand, means that the producer need not incur a cost to produce nothing because inaction is possible.

[8] Some graphical illustrations will have negatively sloped marginal costs for low output levels, but they eventually turn upwards.

A brief comment on the notion of *fixed cost* is appropriate. The cost function, $c(w, y)$, treats y as fixed. Thus, it is a short-run concept. But, all inputs, x, remain variable. No fixed costs (NFC) signals that when all inputs are variable, the producer can use a zero input vector to produce nothing. This contrasts with situations where the producer is obligated, either by contractual or other means, to purchase a fixed input bundle regardless of the amount produced. In that case, fixed costs are present, NFC does not apply, and we cannot show that $c(w, 0) = 0$.

Figure 9 depicts the cost function for a canonical technology in the two-output case. It emanates from the origin and is nondecreasing and convex in y.

Figure 9: Cost Function for a Canonical Technology



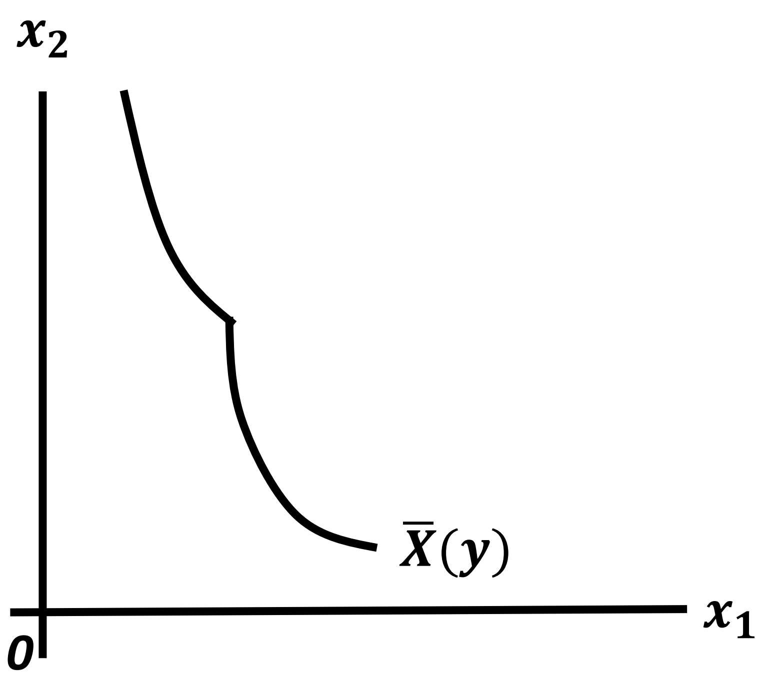
Cost Minimizing Input Demand and Outputs We close this lecture by examining how cost-minimizing demands respond to changes in output, y. General results are scarce. Input demands can respond in different ways to output changes. Sometimes increasing the production of an output means increasing the use of an input while decreasing the use of others. Or it can mean increasing all inputs. Much depends on the setting and the applied circumstances.

Faced with such analytic ambiguity, economists settle for a taxonomy of input behavior. If input k is an increasing function of output m, input k is *progressive* in output m. If input k is a decreasing function of output m, input k is *regressive* in output m. In Figure 6, if the input depicted is progressive in output m, an increase in output m leads that demand curve to shift to the right. Conversely, if it is regressive, it shifts to the left.

Exercises

1. Figure 10 depicts an isoquant for a noncanonical technology. Isolate all the points that could be cost minimizing for some combination of strictly positive input prices. Explain your answer.

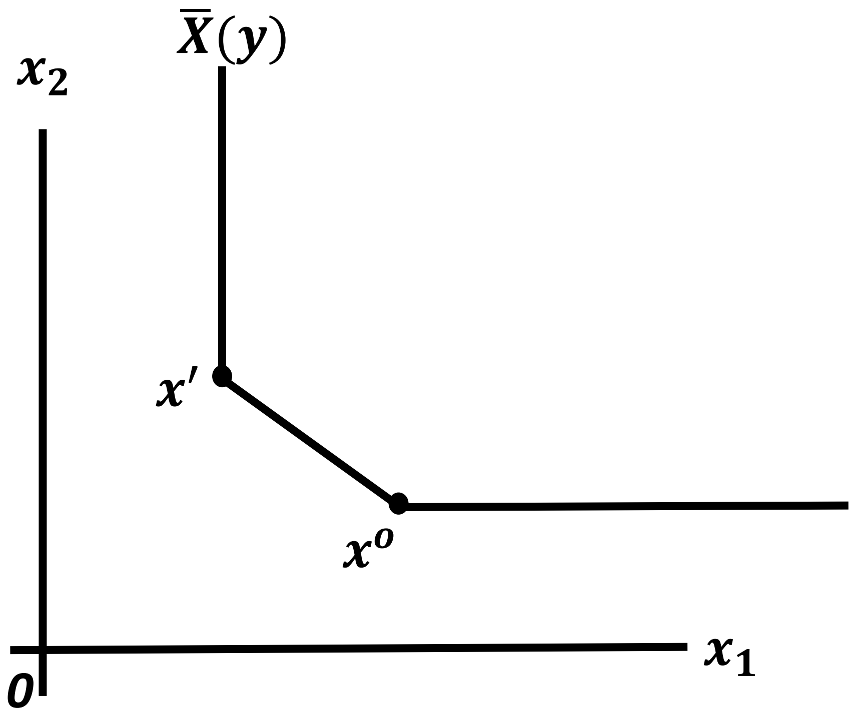
Figure 10: Noncanonical Isoquant



2. Using Figure 4, explain why the producer's best response to price rising from w^* to $\hat{w}$ must involve a minimal cost that satisfies $\hat{w}x^* \geq c(\hat{w}, y)$.

3. Figure 11, which is borrowed from Lecture 4, depicts an isoquant for a canonical technology. Describe how the cost minimizing demands for both inputs change as the real price of input 1, $\frac{w_1}{w_2}$, varies over the interval $0 < \frac{w_1}{w_2} < \infty$. Illustrate in the (x_1, w_1) plane the demand curve for input 1 that results when $w_2 = 1$.

Figure 11: Deriving Input Demands



4. Show that cyclical monotonicity of $c(w, y)$ in w (see footnote 5),

$$w^1 \left(x\left(w^1, y\right) - x\left(w^2, y\right)\right) + w^2 \left(x\left(w^2, y\right) - x\left(w^3, y\right)\right) + ... + w^K \left(x\left(w^K, y\right) - x\left(w^1, y\right)\right) \leq 0$$

implies that

$$\left(w_k^o - w_k'\right)\left(x_k\left(w^o, y\right) - x_k\left(w', y\right)\right) \leq 0$$

if $w_j^o = w_j'$ for $j \neq k$.

5. Explain why convexity of T ensures that marginal cost is a nondecreasing function of output and illustrate it with a picture.

6. Two chicken farmers use the same production process (that is, they both face the same technology) that employs chicken feed and hired labor to produce chickens. The farmers live in different states (Arkansas and Maryland) and therefore face different input and output prices, but it turns out for contractual reasons that they both want to produce the same number of chickens. The price of chicken feed in Arkansas is \$2.00 a pound while the price of hired labor is \$5.00 an hour. In Maryland, the high-cost state, the price of chicken feed is \$4.00 a pound while the price of hired labor is \$10.00 per hour. Decide which region has higher cost, which region employs the most labor, and which region employs the most chicken feed. Explain all of your answers.

7*. The *isocost set* (as distinct from the isocost line) is defined as the set of outputs that give a level of cost no greater than c for input prices w. Its mathematical definition is

$$Y^*(c, w) \equiv \{y : c(w, y) \leq c\}$$

Draw a picture in (y_1, y_2) space illustrating $Y^*(c, w)$ for a two-output canonical technology.

8* Prove that if T is a convex set, $Y^*(c, w)$ is a convex set.

9*. Prove that $Y^*(c, w)$ being a convex set *does not imply* that T is convex.

Lectures on Neoclassical Production Economics[1]

Lecture 7: The Revenue Function

Robert G Chambers [2]

January 7, 2024

[2]University of Maryland

Lecture Introduction

This lecture's subject is the revenue function defined in Lecture 5. It is the money-metric generalization of the single-output production function that solves the problem of adding unlike output quantities by converting them into value terms. The analytic approach echoes that used in Lectures 5 and 6. Now, however, we seek the highest isorevenue line (hyperplane), $\bar{\mathcal{R}}(r,p)$, consistent with a given output set, $Y(x)$. As Figure 1 illustrates, the revenue maximizing solution, denoted by $y(p,x)$, occurs where an isorevenue line (hyperplane) supports the output set from above, so that

$$\bar{\mathcal{R}}(py(p,x),p) \cap Y(x) \neq \emptyset \text{ and} \tag{1}$$

$$Y(x) \subset \mathcal{R}^{-}(py(p,x),x) \tag{2}$$

characterizes the solution.

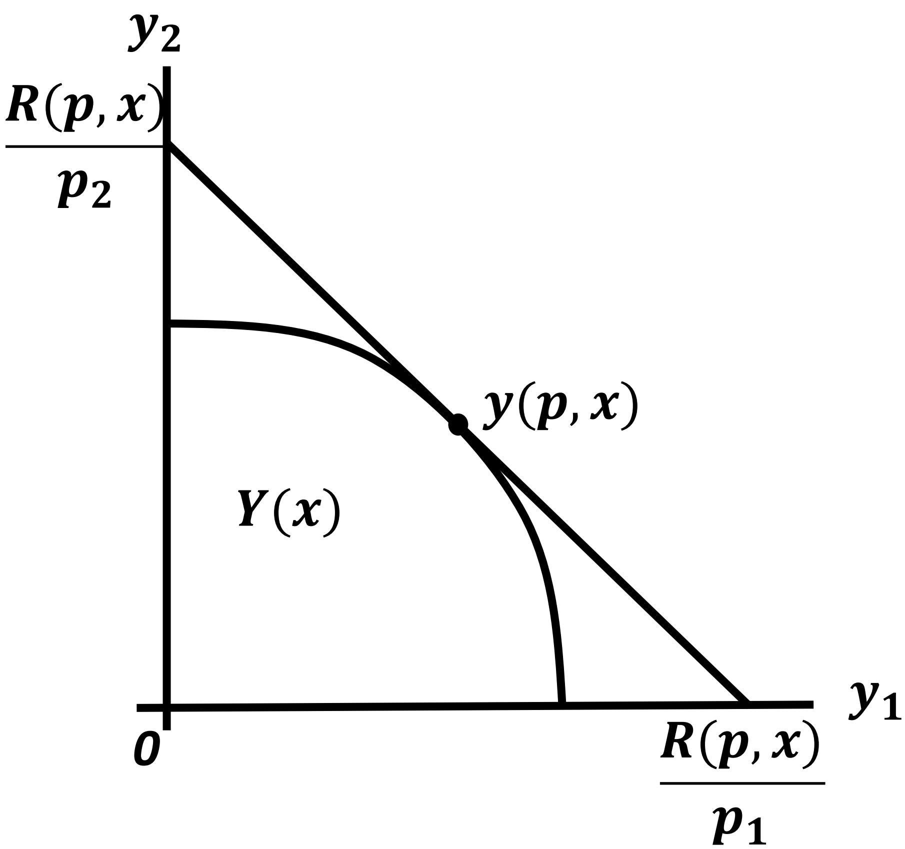
Lecture 6 showed that the behavior of the cost function as a function of input prices only depends upon a mild restriction on the technology. Moreover, the properties of $c(w,y)$ in w characterize the behavior of cost-minimizing behavior input demands in w. The parallel result for the revenue function is that its behavior as a function of output prices only requires a mild (but slightly stronger) restriction upon the technology and that it characterizes the behavior of revenue maximizing supplies in output prices.

We proceed as follows. First, we examine the properties of the revenue function in output prices. Then we examine the consequences of FDI, convexity, NFC, and NFL for $R(p,x)$. Because the argument parallels those Lectures 5 and 6, we adopt a terser style to avoid repeating basic principles. Those readers who are just "dropping in" to this lecture may benefit from reviewing Lectures 5 and 6.

The Revenue Function Defined and its Properties in Output Prices

Definition 1. The revenue function, $R(p,x)$, for output prices p and input bundle x is the maximal revenue that can be realized from the output set for x, $Y(x)$.

Figure 1: Revenue Maximization



We recall Chapter 5's mathematical definition of $R(p,x)$.[1]

$$\begin{aligned} R(p,x) &\equiv \max_y \{py : (x,y) \in T\} \\ &= \max_y \{px : y \in Y(x)\}. \end{aligned}$$

The revenue function is well defined if the output set is nonempty, $Y(x) \neq \emptyset$, and "bounded".[2] We first state mathematical versions of $R(p,x)$'s properties and then follow with a discussion of each.

The Revenue Function as a Function of Output Prices

- $R(p,x) \geq 0$ (nonnegative)
- $R(p',x) \geq R(p,x)$ for $p' \geq p$ (nondecreasing in p)
- $R(\mu p,x) = \mu R(p,x)$ for $\mu > 0$ (positively homogeneous in p)
- $R(\lambda p^o + (1-\lambda)p^*, x) \leq \lambda R(p^o,x) + (1-\lambda) R(p^*,x)$ $\lambda \in (0,1)$ (convex in p)

Nonnegative The revenue function is nonnegative because outputs, y, are nonnegative and output prices are positive. Hence, revenue as their product cannot be negative.

Nondecreasing in p Let y^* be revenue maximizing for p. Because $Y(x)$ does not depend upon p, y^* remains producible for $p' \geq p$. But $p' \geq p$ implies that the revenue for the "do-nothing" response satisfies

$$p'y^* \geq py^* = R(p,x),$$

so that $R(p',x) \geq R(p,x)$. When depicted in (p,r) space, the graph of $R(p,x)$ has a nonnegative slope.

[1] The mathematical name of the revenue function is the (upper) *support function* for $Y(x)$.

[2] Boundedness is a technical requirement that ensures that there are no infinitely large feasible output bundles.

Positively homogeneous in p Inspecting Figure 1 while recalling that the isorevenue line's slope only depends upon real prices implies that a nominal price change leaves that solution unchanged. Once again, 'only real prices matter' is the core principle. Producers respond to real but not nominal price changes. The mathematical argument follows the same method used to demonstrate the positive homogeneity of $c(w,y)$ in w.

$$\begin{aligned}
R(\mu p, x) &= \max_y \{\mu p y : y \in Y(x)\} \\
&= \mu \max_y \{py : y \in Y(x)\} \\
&= \mu R(p, x)
\end{aligned}$$

for $\mu > 0$. Here the key idea is that making μpy as large as possible reduces to making py as large as possible if we hold μ fixed. Thus, if y^*, for example, solves

$$\max_y \{py : y \in Y(x)\},$$

it must also solve

$$\max_y \{\mu py : y \in Y(x)\}.$$

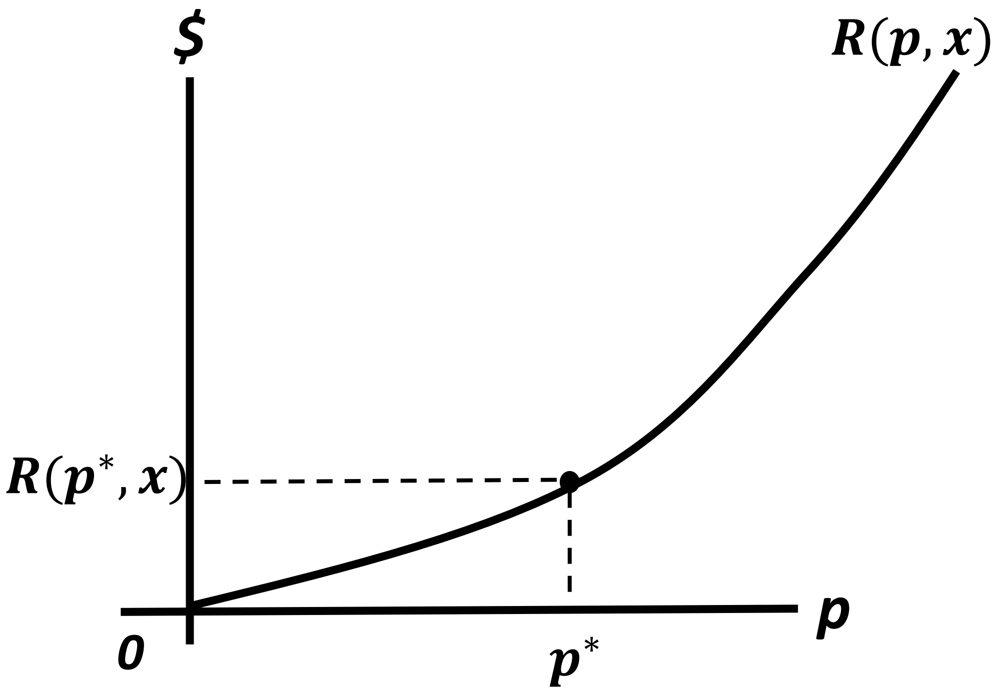
Convex in p A nondecreasing, convex function is visualized as one that has a positive and increasing slope. Figure 2 illustrates. To show the reasoning: pick an arbitrary output-price vector, p^*, in Figure 2 and label its revenue maximizing output bundle y^*, so that $R(p^*, x) = p^* y^*$. Let the output price rise to $p^o > p^*$. (Exercise 2 asks you to developed the associated geometric argument.[3]) How do you illustrate the associated change in revenue if the producer keeps production at y^*? Is that a feasible response? How does it compare to an optimal response? Answering these questions confirms that the convex shape in Figure 2 is not only possible, but expected, if producers can adjust (transform) their output production in response to changes in p.

For a more formal argument, let

$$\hat{p} \equiv \lambda p^o + (1 - \lambda) p^*$$

[3]You may wish to consult Figures 3 and 4 and the surrounding discussion in Chapter 6 in fleshing out the elaboration.

Figure 2: Convex Revenue Function



for $\lambda \in (0,1)$ and $\hat{y} \in Y(x)$ denote a revenue maximizing solution for $\hat{p}$. By definition

$$\begin{aligned} p^o \hat{y} &\leq R(p^o, x) \\ p^* \hat{y} &\leq R(p^*, x). \end{aligned}$$

Multiplying the first by λ, the second by $1-\lambda$, adding, and rerranging gives

$$R(\lambda p^o + (1-\lambda) p^*, x) \leq \lambda R(p^o, x) + (1-\lambda) R(p^*, x),$$

for $\lambda \in (0,1)$ as desired because

$$R(\lambda p^o + (1-\lambda) p^*, x) = (\lambda p^o + (1-\lambda) p^*) \hat{y}.$$

Revenue Maximizing Supplies

Definition 2. A revenue maximizing supply vector, $y(p,x)$, for prices p and input bundle x is any solution to the revenue maximization problem for p and x.

In mathematical form, that definition becomes

$$y(p,x) \in \underset{y}{\operatorname{argmax}} \{py : y \in Y(x)\}.$$

Our investigation of $y(p,x)$'s behavior starts with noting that each such supply vector, $y(p^*,x)$, defines a linear function of output prices p, $py(p^*,x) \equiv \sum_{m=1}^{M} p_m y_m(p^*,x)$, that satisfies

$$p^o y(p^*,x) \leq R(p^o,x)$$

and

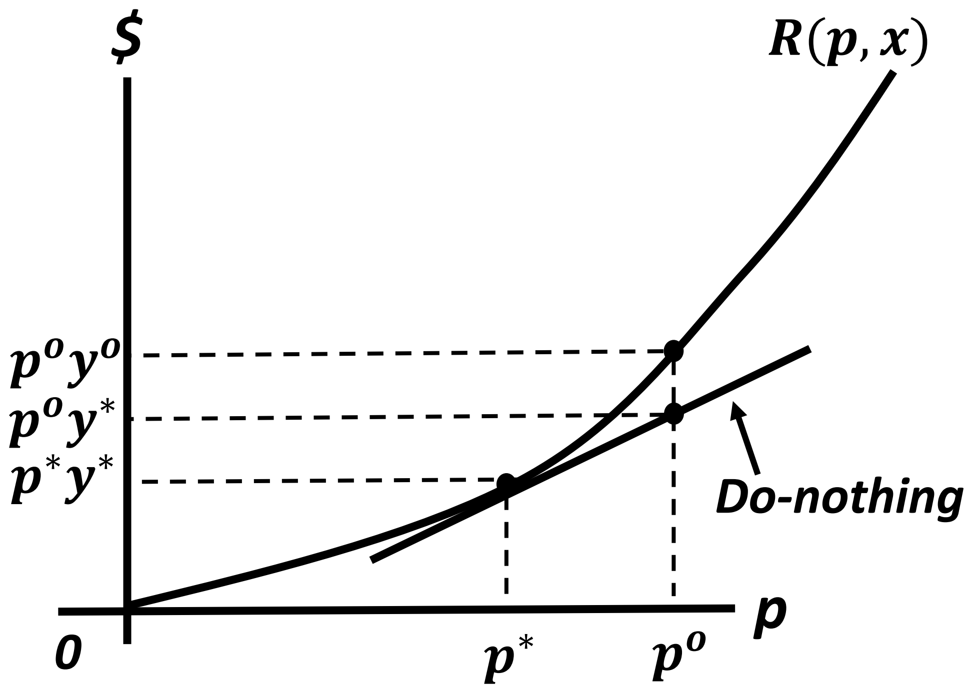
$$p^* y(p^*,x) = R(p^*,x),$$

for all p^* and p^o. Rephrased in geometric terms: the linear function of p^o, $p^o y(p^*,x)$, with normal $y(p^*,x)$ defines a *lower bound* for the revenue function, $R(p^o,x)$, that equals it when $p^o = p^*$. Subtracting the equality from the inequality gives:

$$(p^o - p^*) y(p^*,x) \leq R(p^o,x) - R(p^*,x), \tag{3}$$

for all p^o as a requirement for $y(p,x)$ to be revenue-maximizing.[4] Figure 3 illustrates. The graph of the linear function, $py(p^*,x)$, is illustrated by the line segment passing through

Figure 3: McFadden's Lemma



(p^*, p^*y^*) and $(p^o, p^o y^*)$, where now $y^* = y(p^*, x)$. The economic implication is that the "do-nothing" response of keeping output at $y^* = y(p^*, x)$ as output price changes from p^* to p^o never generates more revenue than the optimal response. Expression (3) is the analogue of Shephard's Lemma for the revenue function.

McFadden's Lemma For all p^* and x, any solution to the revenue-maximization problem, $y(p^*, x)$ must be the normal for a hyperplane tangent to the graph of $R(p, x)$ at $(p^*, R(p^*, x))$.[5]

McFadden's Lemma links revenue-maximizing supplies, $y(p, x)$, to the slope of the revenue function in p. That implies that the behavior of $R(p, x)$ in p will characterize the behavior of $y(p, x)$. For example, we have argued that the revenue-function is positively homogeneous in p, or that 'only real price matters' in determining revenue-maximizing supplies. Let's confirm that argument using McFadden's Lemma and the properties of $R(p, x)$.

Following an argument developed in Chapter 6, we note that if $y(p^*, x)$ satisfies expression (3), positive homogeneity of $R(p, x)$ in p ensures that it also satisfies

$$(\mu p^o - \mu p^*) y(p^*, x) \leq R(\mu p^o, x) - R(\mu p^*, x) \tag{4}$$

for all μp^o and μp^* because positive homogeneity allows expression (4) to be rewritten as

$$\mu (p^o - p^*) y(p^*, x) \leq \mu (R(p^o, x) - R(p^*, x)).$$

And dividing both sides by $\mu > 0$ gives expression (3). Thus, expressions (3) and (4) are equivalent. Applying McFadden's Lemma we conclude, therefore, that

$$y(\mu p, x) = y(p, x) \quad for \;\; \mu > 0.$$

Revenue maximizing supplies are *homogeneous of degree zero* in p, and nominal prices changes do not affect them.

Figure 1 suggests that increasing the price of output 1, p_1, makes the slope of the isorevenue line steeper (from the perspective of the vertical axis). That suggests that the producer

[4]Expression (3) says that the revenue maximizing supply vector is a *subgradient* of $R(p, x)$ at p^*.

[5]Alternatively: Any solution to the revenue-maximization problem at (p^*, x) must be a subgradient for $R(p, x)$ at $R(p^*, x)$.

responds to an increase in p_1 by moving along the transformation curve to a point where more of output 1 is produced and less of output 2. Similarly, Figure 3 shows that the slope of $R(p, x)$ is steeper at p^o than at p^*, which when coupled with McFadden's Lemma also suggests that output responds positively to an increased price.

We confirm these intuitive arguments by using the two inequalities

$$\begin{aligned} p^o y(p^*, x) &\leq p^o y(p^o, x) \\ p^* y(p^o, x) &\leq p^* y(p^*, x) \end{aligned}$$

that are a necessary consequence of revenue maximization for all p^o and p^*. (They say that revenue maximizing supplies generate more revenue at the prices for which they are revenue maximizing than revenue maximizing supplies for other prices.) Adding them together and rearranging gives[6]

$$(p^o - p^*)(y(p^o, x) - y(p^*, x)) \geq 0, \tag{5}$$

or in expanded form

$$\sum_{m=1}^{M} (p_m^o - p_m^*)(y_m(p^o, x) - y_m(p^*, x)) \geq 0.$$

Revenue maximizing supplies and output-prices tend to vary together. Taking $p_k^o = p_k^*$ for all $k \neq j$ gives

$$\left(p_j^o - p_j^*\right)(y_j(p^o, x) - y_j(p^*, x)) \geq 0,$$

which confirms that all revenue-maximizing supplies are upward sloping (more precisely not downward sloping) in their own prices. Figure 4 illustrates.

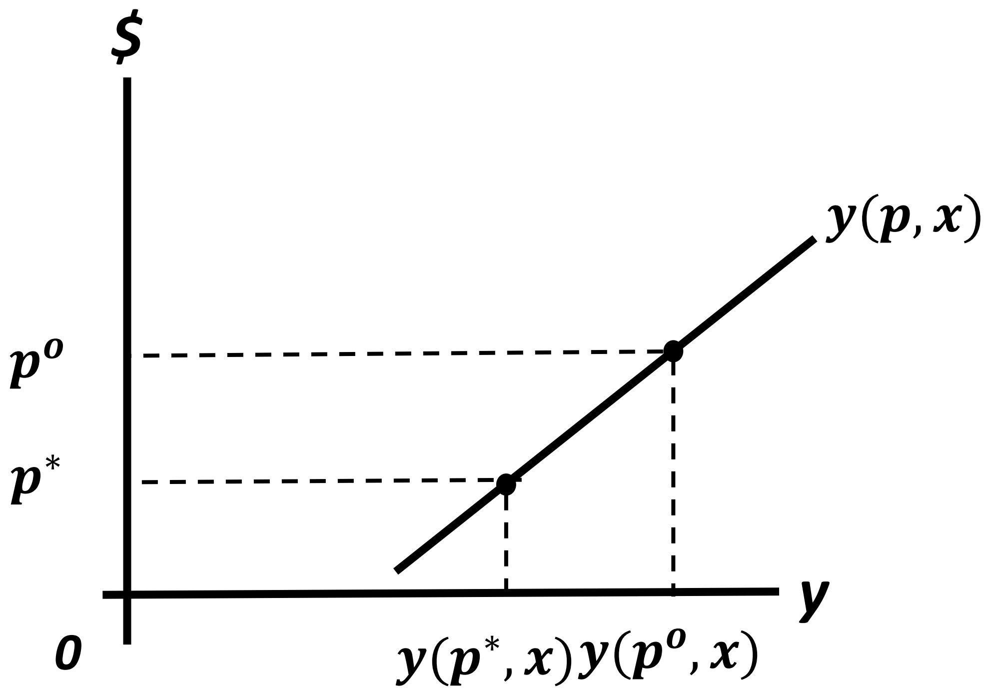
From the discussion of Figure 1, we anticipate that an increase in the price of one output tends to suppress production of other commodities. Using a bit more jargon, we might therefore conclude that *cross price adjustments* of revenue maximizing supplies tend to be

[6] Following the line of argument in footnote 5 of Lecture 6, we can show that expression (5) is a special case of revenue-maximizing supplies satisfying the following *cyclical monotonicity* property in p:

$$p^1\left(y\left(p^1, x\right) - y\left(p^2, x\right)\right) + p^2\left(y\left(p^2, x\right) - y\left(p^3, x\right)\right) + \dots + p^K\left(y\left(p^K, x\right) - y\left(p^1, x\right)\right) \geq 0.$$

Together with zero homogeneity, cyclical monotonicity exhaustively characterizes the behavior of revenue maximizing supplies in p.

Figure 4: Revenue Maximizing Supply



negative (more precisely nonpositive). If FDO applies, and there are only two outputs, this is indeed true. More generally, it's not.

Consider, for example, a farmer who produces wheat and beef for the local market. Because a necessary by-product of beef production is hide production (which can be used in leather production), one expects that an increase in the price of beef would lead to an increase in hide production even if the price of hides remained constant. In fact, one would not be surprised if an increase in the price of beef led to an increase in hide production *even if the price of hides fell.* Such things can and do happen.

Outputs whose supply increases as cross price p_k increases are *complements* for output k. *Substitutes* for k are outputs whose supply decreases as p_k falls. Referring to Figure 4, let the supply illustrated there be for output $m \neq k$ so that the price illustrated on the vertical axis will be p_m. If output m is a substitute for output k, an increase in p_k will cause the demand curve to shift to the left so that less of output m is supplied at all levels of p_m. On the other hand, if output m is a complement for output k, an increase in p_k will cause the supply curve for output m to shift to the right at the production of m increases with the production of k.

Revenue Function as a Function of Inputs, x

How $R(p, x)$ responds to changes in x depends upon $Y(x)$. Because $R(p, x)$ is the money-metric generalization of the production function, $f(x)$, that accommodates multiple outputs (for the single-output case $R(p, x) = pf(x)$), its behavior in x has natural parallels in the behavior of $f(x)$. We first list the implications for $R(p, x)$ of restrictions imposed on $Y(x)$ and then discuss them.

Properties of $R(p, x)$ in x

- $FDI \Rightarrow R(p, x') \geq R(p, x)$ for $x' \geq x$ (nondecreasing in x)
- Convexity of $T \;\Rightarrow\; R(p, \lambda x^o + (1-\lambda) x^*) \geq \lambda R(p, x^o) + (1-\lambda) R(p, x^*)$ for $\lambda \in (0, 1)$ (concave in input)

- *NFL and NFC* $\Rightarrow R(p,0)=0$.

Nondecreasing in x Free disposability of inputs ensures that maximal revenue cannot decrease as a result of an increase in the input bundle, x. The demonstration follows the "do nothing" principle. Because $y(p,x) \in Y(x)$, FDI ensures that it also belongs to $Y(x')$ for $x' \geq x$. Therefore, the producer can respond to an input increase by producing the same output bundle. And because that choice results in no revenue change, the best response cannot decrease revenue.

When $R(p,x)$ is graphed as a function of x, nondecreasing in x rules out a negatively sloped curve. The parallel property for the production function was called a positive (non-negative) marginal productivity. Here the parallel concept is *marginal revenue*. If there is only a single input, *marginal revenue* is defined as change in revenue divided by the change in the input, $\frac{\Delta R(p,x)}{\Delta x}$, or as the slope of the line segment tangent to the graph of $R(p,x)$ at $(x, R(p,x))$. In the multiple input case, the marginal revenues form the normal to the hyperplane tangent to $R(p,x)$ at $(x, R(p,x))$. They are all nonnegative under FDI.

Concavity as a function of x Concavity of $R(p,x)$ as a function of x echoes the concavity of the production function. It manifest the principle of diminishing returns and implies that successive marginal input increases evoke decreasing (more precisely nonincreasing) marginal revenue increases. In more casual terms, convexity of T implies decreasing marginal revenue for each of the inputs. This evokes a revenue function with the same basic shape as the production function presented in Figure 1 of Lecture 3. The only change is that the vertical axis is denominated in nominal value (dollar) terms rather than in terms of output.

The more formal demonstration follows a now well-trod path. For x^o and x^* find $y(p,x^o)$ and $y(p,x^*)$. Convexity of T ensures that $\lambda y(p,x^o) + (1-\lambda) y(p,x^*) \in Y(\lambda x^o + (1-\lambda) x^*)$ for $\lambda \in (0,1)$. But that requires that

$$R(p, \lambda x^o + (1-\lambda) x^*) \geq p(\lambda y(p,x^o) + (1-\lambda) y(p,x^*)),$$

and the desired conclusion follows immediately from the definition of $y(p,x)$ as revenue maximizing.

NFL and NFC No free lunch ensures that a positive amount of any output cannot be produced by a zero input bundle. No fixed cost, on the other hand, means that 0 is a feasible output choice for a 0 input. Together, they imply that $R(p, 0) = 0$.

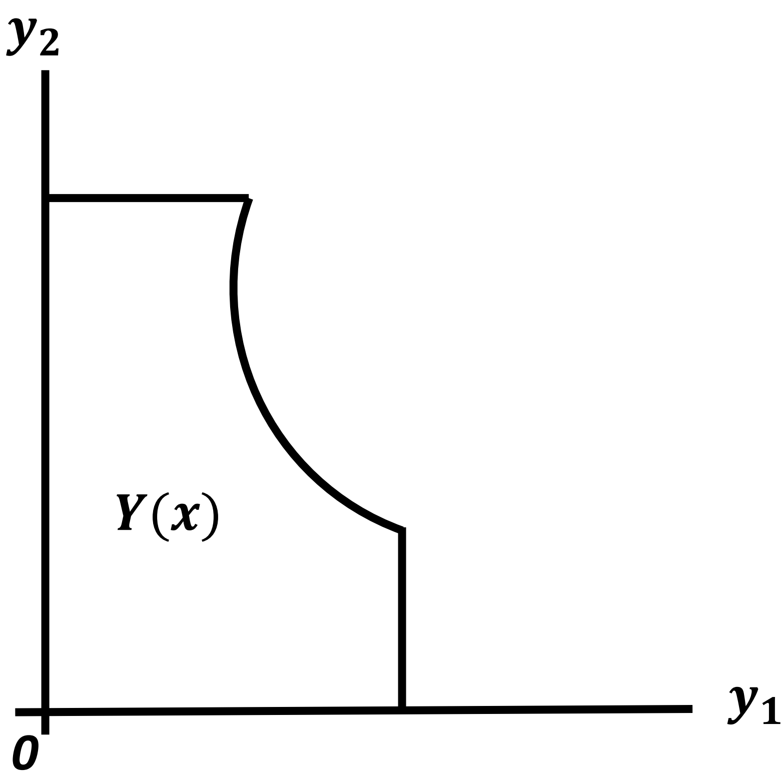
Revenue Maximizing Supply and Input Use As a general principle, we expect an increase in the application of inputs to lead producers to increase their outputs. But it is also easy to imagine instances where increasing some inputs will cause certain revenue maximizing outputs to increase and others to decrease. Let's consider a farmer producing a livestock output and a crop output from a bundle of inputs that consists of land, fertilizer, livestock feed, and labor. If livestock feed increases, one might envision the farmer diverting some labor from crop production to increased feed handling and distribution. The result would be increased livestock production but decreased crop production. Similarly, an increase in fertilizer availability might prompt the farmer to divert labor from livestock to crop-production activities.

If increasing input k is associated with a increase in the revenue maximizing amount of output m, output m is *normal* or *progressive* in input k. If increasing input k is associated with a decrease in the revenue maximizing amount of output m, output m is *regressive* in input k. If output m is normal in input k, an increase in input k causes the supply curve for output m to shift to the right, leading to more output supplied for all levels of p_m. If output m is regressive in input k, the same change in x_k causes the supply curve to shift to the left.

Exercises

1. Figure 5 depicts an output set for a technology. As $\frac{p_1}{p_2}$ varies over the interval $0 < \frac{p_1}{p_2} < \infty$, identify the revenue-maximizing supplies for both outputs. Graph the resulting revenue-maximizing supply for output 2 in (y_2, p_2) space.

Figure 5: Solving the revenue maximization problem



2. Develop a geometric illustration of the demonstration that $R(p, x)$ is convex in p.

3. Develop a geometric illustration of a revenue function in (x, r) space for a canonical technology with two outputs and one input.

4. Consider a farmer who is producing both corn and wheat using two inputs, fertilizer and land. Also assume that the farmer has a given amount of fertilizer and land is trying to make as much profit from them as possible. The price of corn is initially \$4/bushel and the

price of wheat is initially \$7/bushel. Explain how the farmer decides how much wheat to produce and how much corn to produce. What happens to the supply of wheat, the supply of corn, and your revenue if the price of wheat fell from \$7 to \$5.

5.* Consider three output-price vectors, p^1, p^2, and p^3 and their respective revenue-maximizing supplies. By the definitions of the revenue function and revenue-maximizing supplies:

$$\begin{aligned}
p^1 y\left(p^1, x\right) &\geq p^1 y\left(p^2, x\right) \\
p^2 y\left(p^2, x\right) &\geq p^2 y\left(p^3, x\right) \\
p^3 y\left(p^3, x\right) &\geq p^3 y\left(p^1, x\right).
\end{aligned}$$

Adding these three expressions together and rearranging gives

$$p^1\left(y\left(p^1, x\right) - y\left(p^2, x\right)\right) + p^2\left(y\left(p^2, x\right) - y\left(p^3, x\right)\right) + p^3\left(y\left(p^3, x\right) - y\left(p^1, x\right)\right) \geq 0.$$

Show that this last expression implies

$$\left(p^1 - p^3\right)\left(y\left(p^1, x\right) - y\left(p^3, x\right)\right) \geq 0,$$

and explain its economic meaning intuitively.

Lectures on Neoclassical Production Economics[1]

Lecture 8: The Profit Function

Robert G Chambers[2]

January 8, 2024

[2]University of Maryland

Lecture Introduction

We introduced the profit function, $\pi(p, w)$, in Chapter 5 and discussed it for the single-input, single-output technology. This lecture examines it and its associated profit maximizing supplies and input demands for general technologies. The analytic principles remain unchanged, the producer seeks a point where an isoprofit line (hyperplane) supports T from above. Thus, the intuitive arguments and visual illustrations often parallel ones used in discussing $c(w, y)$ and $R(p, x)$. That often makes for a terser argument than adopted in Chapters 5-7.

We start the chapter by discussing some nuances associated with using a maximal value function to characterize producer behavior that are easy to overlook. Then we turn to an examination of the properties of the profit function defined in terms of the technology, T:

$$\pi(p, w) \equiv \max_{x,y} \{py : (x, y) \in T\},$$

and we close the Lecture by discussing the two-stage versions of the profit maximization problem:

$$\pi(p, w) \equiv \max_{x} \{R(p, x) - wx\}$$

and

$$\pi(p, w) \equiv \max_{y} \{py - c(w, y)\}.$$

Maxima and Behavior

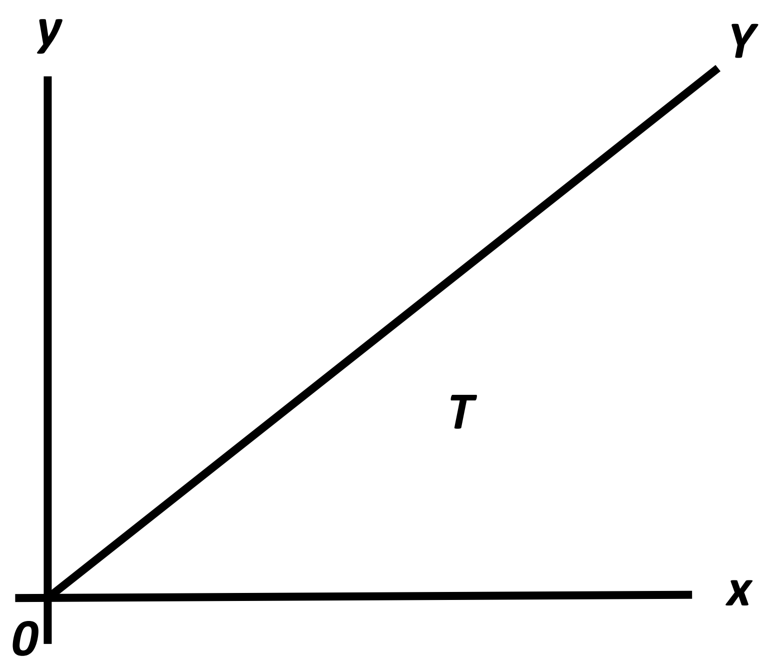
Figure 1 depicts T as everything on or below the ray from the origin labelled 0Y. You should verify that this representation corresponds to a canonical technology. But it's important to note that the linear nature of $T's$ upper bound also ensures that T satisfies:[1]

- **Constant Returns to Scale** $(x, y) \in T \Rightarrow (\mu x, \mu y) \quad \mu > 0$.

Suppose that a producer finds a profitable (x, y) bundle. Can the producer produce $2(x, y)$ and double profits, $3(x, y)$ and triple profits, and so on? The answer is yes for a constant-returns-to-scale (more simply, constant-returns) T.

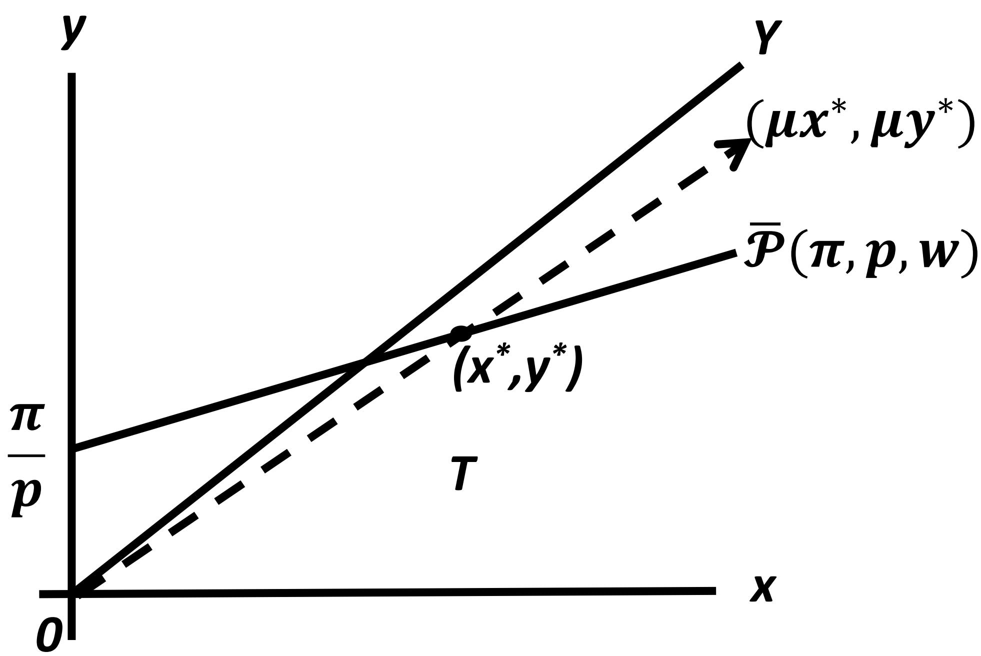
[1] A constant-returns-to-scale technology set, T is a special case of a mathematical object called a *cone*. For that reason, some writers refer to constant-returns-to-scale technologies as being *conical*. By definition, $K \subset \mathbb{R}^J$ is a *cone* if $k \in K \Rightarrow \lambda k \in K$ for all $\lambda > 0$.

Figure 1: Constant Returns Technology



Interpreting the profit function and the profit maximizing solution for a constant-returns technology is subtle.[2] Figure 2 superimposes an isoprofit line, $\bar{\mathcal{P}}(\pi, p, w)$, on the technology in Figure 1. Any point on $\bar{\mathcal{P}}(\pi, p, w)$ that lies inside T is technically feasible and yields a profit of $\pi > 0$. (Note the vertical intercept of the isoprofit line is positive.) For example, take the point labelled (x^*, y^*). Because the technology satisfies constant returns, $(\mu x^*, \mu y^*) \in T$

Figure 2: Infinite Profit and Constant Returns



even as μ becomes infinitely large. But if $\pi = py^* - wx^* > 0$, $\mu\pi = p\mu y^* - w\mu x^*$ goes to infinity as μ is made arbitrarily large. Therefore, if $\pi > 0$ is available for a constant returns

[2] Our discussion, which focuses on producer decisions, rarely treats issues associated with economies of scale or economies of scope. However, these matters assume central importance in studying the nature of competitive equilibrium with free entry and exit of both producers and consumers. And there constant-returns technologies represent an important benchmark in studying firm behavior.

to scale technology, so too is an infinite profit and $\pi(p, w) = \infty$.[3]

On the other hand, if input and output prices are such that the isocost line, $\bar{\mathcal{P}}(\pi, p, w)$, is more steeply sloped (not drawn) than 0Y, maximal profit occurs at the origin (see Exercise 1). In this case, firms do not operate and profit is zero. Zero profit also occurs, regardless of how much is produced, in the knife-edge case where relative prices define isoprofit lines having the same slope as the boundary of the constant-returns T. Thus, for the constant returns technology, $\pi(p, w)$ always assumes one of two bounding values either 0 or ∞. [4]

In figure 3, the technology set is depicted as everything on or below the curve AY. This T violates no fixed cost, NFC. Nevertheless, profit is maximized where an isoprofit line, $\bar{\mathcal{P}}(\pi, p, w)$, supports T from above. As drawn, however, $\frac{\pi}{p} < 0$, maximal profit is negative for the relative price that defines the slope of $\bar{\mathcal{P}}(\pi, p, w)$.

Is it plausible to believe that a producer would tolerate a negative profit? We introduced the technology, T, as a constraint on *production choices*. But we have not discussed *what causes a decisionmaker to be a producer*. Is the decisionmaker free to pursue alternative employment? In a free-market setting, the answer is presumably yes. And decisionmakers are free to choose between acting as producers and earning $\pi(p, w)$ or pursuing alternative employment. Economists call the *best-possible remuneration from alternative employment* the decisionmaker's *opportunity cost* of being a producer. To keep things simple, suppose that the opportunity cost is 0. Then the decisionmaker's employment choice is driven by

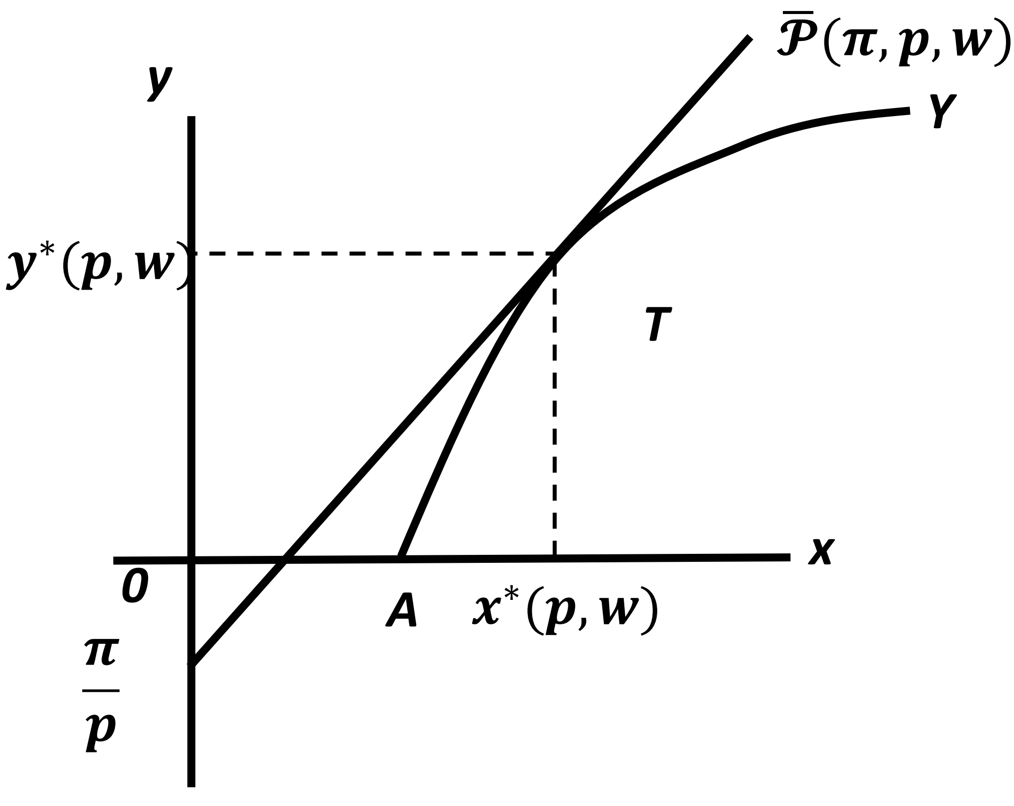
$$\max\{0, \pi(p, w)\},$$

where $0 > \pi(p, w)$ signals that the decisionmaker is better off pursuing alternative employment. Figure 3 depicts such a case.

[3] Infinity is not a real number, so what this means is that profit has an infinite limit in such cases. In any case, the intuition is apparent. Constant returns allows producers to make unbounded profits. Hence, it's hard to believe that technologies exist for which constant returns applies globally.

[4] Why market prices should fall in any of these regions is precisely the "stuff" of competitive market analysis and lies beyond the scope of our discussion.

Figure 3: Should the Decisionmake be a Producer?



Properties of the Profit Function

Definition 1. The profit function, $\pi(p,w)$, gives the maximal feasible profit at output prices p and input prices w.

The mathematical definition is:[5]

$$\pi(p,w) \equiv \max_{x,y} \{py - wx : (x,y) \in T\}.$$

We state $\pi(p,w)$'s properties and then discuss them.

Properties of the Profit Function

- $p' \geq p \Rightarrow \pi(p',w) \geq \pi(p,w)$ and $w' \geq w \Rightarrow \pi(p,w') \leq \pi(p,w)$ (nondecreasing in p and nonincreasing in w)
- $\pi(\mu p, \mu w) = \mu\pi(p,w) \quad \mu > 0$ (positively homogeneous)
- $\pi(\lambda p^o + (1-\lambda)p^*, \lambda w^o + (1-\lambda)w^*) \leq \lambda\pi(p^o,w^o) + (1-\lambda)\pi(p^*,w^*) \quad \lambda \in (0,1)$ (convex)

Nondecreasing in p and nonincreasing in w Let (x^*,y^*) be profit maximizing for (p,w). By responding to a price rise of p to p' by "doing nothing"', the producer realizes a profit of

$$p'y^* - wx^* \geq py^* - wx^* = \pi(p,w).$$

The producer's "best response" to the price change necessarily dominates the "do-nothing response". Hence, $\pi(p',w) \geq \pi(p,w)$ for $p' \geq p$. By responding to an input price decrease from w' to w by "doing nothing" the producer realizes a profit of

$$py^* - wx^* \geq py^* - w'x^* = \pi(p,w').$$

The producer's "best response" necessarily dominates the "do-nothing" response, whence $\pi(p,w') \leq \pi(p,w)$ for $w' \geq w$.

When graphed in the (p,π) plane, $\pi(p,w)$ has a nonnegative slope and when graphed in the (w,π) plane a nonpositive slope.

[5] The math literature refers to $\pi(p,w)$ as a *support function* for the set T in terms of $(p,-w)$.

Positively homogeneous in prices The underlying principle is that "only real prices matter". Visually, the principle is captured by remembering that nominal price changes do not affect an isoprofit line's slope, $\bar{\mathcal{P}}(\pi, p, w)$. (See Lecture 5, Figure 3). Hence, where an isoprofit line (hyperplane) supports T is not affected by nominal price changes.

The formal demonstration follows the same line of argument used in discussing $c(w, y)$ and $R(p, x)$:

$$\begin{aligned}
\pi(\mu p, \mu w) &= \max_{x,y}\{\mu p y - \mu w x : (x, y) \in T\} \\
&= \max_{x,y}\{\mu(py - wx) : (x, y) \in T\} \\
&= \mu \max_{x,y}\{py - wx : (x, y) \in T\} \\
&= \mu \pi(p, w) \quad \mu > 0.
\end{aligned}$$

Convex in prices Let (x^*, y^*) be profit maximizing for (p, w). If prices change to (p^o, w^o), the "do-nothing" profit change is

$$(p^o - p)\, y^* - (w^o - w)\, x^*,$$

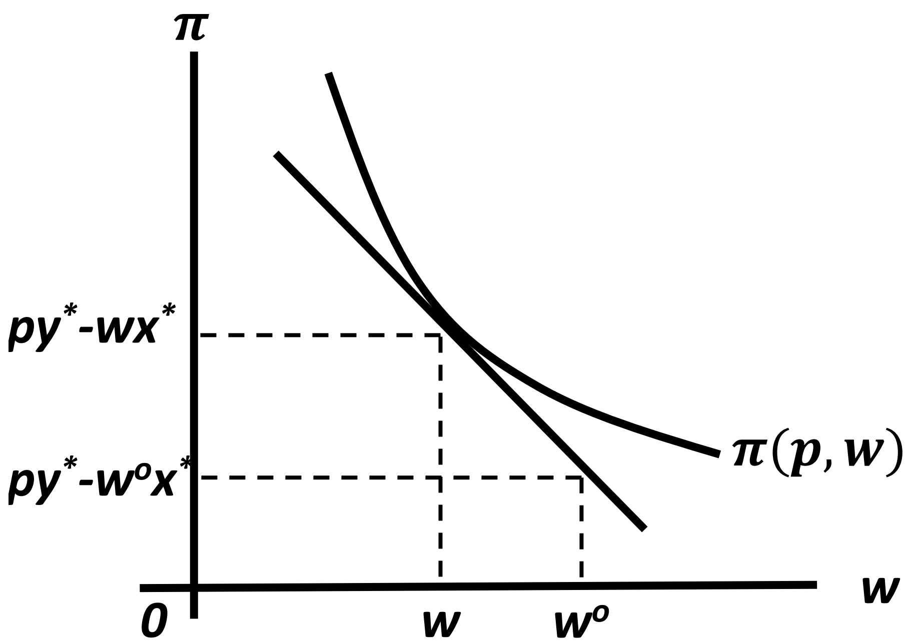
which defines a linear function of the price changes $(p^o - p)$ and $(w^o - w)$. By definition, the producer's best response must at least weakly dominate this "do-nothing"response. Thus, the best-response profits change must "better than linear".

Figure 4 depicts a "better than linear" response for an input-price increase. (Exercise 2 asks you to develop the analogous figures for an output price change.) If the input price rises to w^o from w, and the producer stays at (x^*, y^*), profit falls as costs increase from wx^* to $w^o x^*$. The negatively sloped line segment connecting $(w, py^* - wx^*)$ and $(w^o, py^* - w^o x^*)$ illustrates. (The slope of the line segment is $-x^*$.) The producer's best response must at least weakly dominate the "do-nothing" response, suggesting that $\pi(p, w)$ falls less than $(w^o - w)\, x^*$.

The formal demonstration that $\pi(p, w)$ is convex in prices also follows a familiar path. We need to compare $\pi(p^o, w^o)$, $\pi(p^*, w^*)$, and $\pi(\lambda p^o + (1 - \lambda) p^*, \lambda w^o + (1 - \lambda) w^*)$. Let $(\hat{x}, \hat{y}) \in T$ solve the profit maximization problem for $\lambda p^o + (1 - \lambda) p^*$ and $\lambda w^o + (1 - \lambda) w^*$, so that

$$\pi(\lambda p^o + (1 - \lambda) p^*, \lambda w^o + (1 - \lambda) w^*) = (\lambda p^o + (1 - \lambda) p^*)\, \hat{y} - (\lambda w^o + (1 - \lambda) w^*)\, \hat{x}. \quad (1)$$

Figure 4: Profit Function Convex in Input Prices



The definition of $\pi(p,w)$ ensures that

$$\begin{aligned} p^o\hat{y} - w^o\hat{x} &\leq \pi(p^o, w^o) \\ p^*\hat{y} - w^*\hat{x} &\leq \pi(p^*, w^*) . \end{aligned}$$

That is, the profit from producing $(\hat{x}, \hat{y})$ when prices are (p^o, w^o) and (p^*, w^*), respectively, cannot exceed maximal profit at those prices. Multiplying the first inequality by $\lambda > 0$, the second inequality by $1 - \lambda > 0$, and adding gives

$$(\lambda p^o + (1-\lambda) p^*) \hat{y} - (\lambda w^o + (1-\lambda) w^*) \hat{x} \leq \lambda \pi(p^o, w^o) + (1-\lambda) \pi(p^*, w^*) .$$

We can now rely on expression (1) to replace the left-hand side of this expression to get

$$\pi(\lambda p^o + (1-\lambda) p^*, \lambda w^o + (1-\lambda) w^*) \leq \lambda \pi(p^o, w^o) + (1-\lambda) \pi(p^*, w^*) \quad \lambda \in (0,1) ,$$

which establishes the desired convexity property.

Profit Maximizing Output Supply and Input Demand

We now study the role that the properties of $\pi(p,w)$ play in determining optimal producer behavior. Our starting point is a verbal definition of profit-maximizing input demands and output supply.

Definition 2. Any solution, $(x^*(p,w), y^*(p,w))$, to the profit maximization problem is a profit maximizing input demand and output supply bundle.

In mathematical terms,

$$(x^*(p,w), y^*(p,w)) \in \underset{x,y}{\operatorname{argmax}} \{py - wx : (x,y) \in T\} .$$

Here $\operatorname{argmax}_{x,y} \{\cdot\}$ selects the input and output bundles that give the maximal value in the set $\{\cdot\}$.[6] In Chapters 6 and 7, we learned that optimal solutions to the cost-minimization and revenue-maximization problems are normals for hyperplanes tangent to the graphs of

[6] We intentionally say bundles rather than bundle here because in appropriate circumstances, more than one input-output configuration can achieve the same maximal profit for given (p,w).

$c(w,y)$ and $R(p,x)$, respectively. A bit more informally, cost-minimizing demands are given by the slopes of $c(w,y)$ in w and revenue-maximizing supplies by the slopes of $R(p,x)$ in p. Figure 4, which shows that a line segment with slope $-x^*(p,w)$ supports the graph of $\pi(p,w)$ from below in (w,π) space, suggests a similar phenomenon applies to the profit function, $\pi(p,w)$. (Exercise 3 asks you to to develop and illustrate the analogous property in the output price.) The general result is:[7]

Hotelling's Lemma For all p^* and w^*, $(-x^*(p^*,w^*), y^*(p^*,w^*))$ is the normal for a hyperplane tangent to the graph of $\pi(p,w)$ at $(w^*,p^*,\pi(p^*,w^*))$.

We establish Hotelling's Lemma using arguments that parallel those used for Shephard's Lemma and McFadden's Lemma. Because $(x^*(p^*,w^*), y^*(p^*,w^*)) \in T$, the definition of $\pi(p,w)$ requires that

$$py^*(p^*,w^*) - wx^*(p^*,w^*) \leq \pi(p,w)$$

for all (p,w) and (p^*,w^*), and

$$p^*y^*(p^*,w^*) - w^*x^*(p^*,w^*) = \pi(p^*,w^*).$$

That is, maximal profit for (p,w) is always greater than or equal to profit available from $(x^*(p^*,w^*), y^*(p^*,w^*))$ at (p,w) and equal to it when $(p,w) = (p^*,w^*)$. Subtracting the equality from the inequality gives,

$$(p-p^*)\,y^*(p^*,w^*) - (w-w^*)\,x^*(p^*,w^*) \leq \pi(p,w) - \pi(p^*,w^*), \tag{2}$$

for all (p,w) as a condition that must hold at the profit-maximizing solution.[8] Expression (2) reiterates that doing nothing is dominated profit-wise by the optimal response. Its geometric manifestation is that a hyperplane with normal $(-x^*(p^*,w^*), y^*(p^*,w^*))$ must support the graph of $\pi(p,w)$ from below at $(w^*,p^*,\pi(p^*,w^*))$.

[7]Hotelling's Lemma is often phrased in terms of gradients of the profit function. As with Shephard's Lemma, the gradient version requires imposing further structure upon T. Our statement applies for the general case.

[8]An input-output bundle, $(-x^*(p^*,w^*), y^*(p^*,w^*))$, satisfying expression (2) is called a *subgradient* for the profit function at (w^*,p^*).

"Only real prices matter" in determining cost-minimizing demands and revenue-maximizing supplies. A parallel result applies for profit-maximizing input demands and output supplies. To show this property, we use the positive homogeneity property of the profit function. Multiplying both sides of expression (2) by $\mu > 0$ gives

$$(\mu p - \mu p^*)\, y^*(p^*, w^*) - (\mu w - \mu w^*)\, x^*(p^*, w^*) \leq \mu\left(\pi(p, w) - \pi(p^*, w^*)\right).$$

By invoking positive homogeneity of $\pi(p, w)$, we can rewrite the right-hand side of this expression as $\pi(\mu p, \mu w) - \pi(\mu p^*, \mu w^*)$. Making that substitution, we conclude that

$$(\mu p - \mu p^*)\, y^*(p^*, w^*) - (\mu w - \mu w^*)\, x^*(p^*, w^*) \leq \pi(\mu p, \mu w) - \pi(\mu p^*, \mu w^*)$$

for all (p, w, p^*, w^*) and $\mu > 0$. So, if (2) is satisfied by $(x^*(p^*, w^*), y^*(p^*, w^*))$, it will remain satisfied for all nominal changes of (p, w) and (p^*, w^*). "Only real prices matter" to profit maximizers and

- $(x^*(\mu p, \mu w), y^*(\mu p, \mu w)) = (x^*(p, w), y^*(p, w))$ for $\mu > 0$,

optimal supply and input demand bundles are *homogeneous of degree zero* in (p, w).

Having concluded that nominal price changes do not cause profit-maximizing producers to adjust their inputs and outputs, it's natural to wonder how they respond to real or relative price changes. Shephard's Lemma, when coupled with concavity of $c(w, y)$ in w, taught us that cost-minimizing producers respond to an increase in the price of an input by decreasing use of that input. McFadden's Lemma and convexity of $R(p, x)$ in p taught us that revenue-maximizing producers react to an increase in p_m by increasing their supply of y_m. You are asked in Exercise 4 to draw similar lessons from the convexity of $\pi(p, w)$ and Hotelling's Lemma, while we restrict ourselves to an analysis that only requires a "do-nothing" argument, addition, and a tiny bit of algebra.

Because profit maximizing input and output bundles belong to T,

$$\begin{aligned} p^o y^*(p^o, w^o) - w^o x^*(p^o, w^o) &\geq p^o y^*(\hat{p}, \hat{w}) - w^o x^*(\hat{p}, \hat{w}), \text{ and} \\ \hat{p} y^*(\hat{p}, \hat{w}) - \hat{p} x^*(\hat{p}, \hat{w}) &\geq \hat{p} y^*(p^o, w^o) - \hat{w} x^*(p^o, w^o), \end{aligned}$$

for all (p^o, w^o) and $(\hat{p}, \hat{w})$. The left-hand side of these expressions is, by definition, maximal profit for, respectively, (p^o, w^o) and $(\hat{p}, \hat{w})$. The right-hand side is profit realized from producing the profit maximizing bundle for prices $(\hat{p}, \hat{w})$ and (p^o, w^o), respectively, when (p^o, w^o) and $(\hat{p}, \hat{w})$ prevail. Adding the inequalities and collecting terms gives:

$$(p^o - \hat{p})\left(y^*\left(p^o, w^o\right) - y^*\left(\hat{p}, \hat{w}\right)\right) - \left(w^o - \hat{w}\right)\left(x^*\left(p^o, w^o\right) - x^*\left(\hat{p}, \hat{w}\right)\right) \geq 0. \tag{3}$$

Expression (3) generalizes results obtained for cost-minimizing demands and revenue-maximizing supplies.[9] If $\hat{p} = p^o$ expression (3) implies

$$\left(w^o - \hat{w}\right)\left(x^*\left(p^o, w^o\right) - x^*\left(p^o, \hat{w}\right)\right) \leq 0.$$

Profit maximizing input demands tend to vary inversely with input price changes. Allowing only one input price, say for input k, to change gives

$$\left(w_k^o - \hat{w}_k\right)\left(x_k^*\left(p^o, w^o\right) - x_k^*\left(p^o, \hat{w}\right)\right) \leq 0.$$

Profit maximizing input demands are nonincreasing functions of their own input prices. Input demand functions, when portrayed in the (x_k, w_k) plane, have a nonpositive slope. There are no "Giffen" profit-maximizing input demands.

Exactly parallel arguments holding input prices constant establish that

$$\left(p^o - \hat{p}\right)\left(y^*\left(p^o, w^o\right) - y^*\left(\hat{p}, w^o\right)\right) \geq 0,$$

and

$$\left(p_m^o - \hat{p}_m\right)\left(y_m^*\left(p^o, w^o\right) - y_m^*\left(\hat{p}, w^o\right)\right) \geq 0.$$

Profit maximizing supply curves are nondecreasing functions of their own output prices; supply curves are upward sloping.

[9] Following an argument established in footnote 5 of Lecture 6 and continued in footnote 5 of Lecture 7, we can show that (3) is a special case of the more general *cyclical monotonicity* result for profit-maximizing supplies and input demands:

$$\begin{array}{rcl} p^1\left(y^*\left(p^1, w^1\right) - y^*\left(p^2, w^2\right)\right) & +...+ & p^K\left(y^*\left(p^K, w^K\right) - y^*\left(p^1, w^1\right)\right) \\ -w^1\left(x^*\left(p^1, w^1\right) - x^*\left(p^2, w^2\right)\right) & -...- & w^K\left(x^*\left(p^K, w^K\right) - x^*\left(p^1, w^1\right)\right) \geq 0, \end{array}$$

which together with zero homogeneity exhaustively characterizes profit maximizing supply and input demand behavior.

Two-stage Profit Maximization

Many introductory treatments of profit maximizing behavior limit themselves to the single-output case and postulate the existence of a production function and a cost function. For the production function, $f(x)$, the profit maximization problem is posed as maximizing the difference between the nonlinear function of inputs, $pf(x)$, and the linear function, wx. For the cost function, $c(w,y)$, the equivalent problem is to maximize the difference between the linear function of outputs, py, and the nonlinear cost function, $c(w,y)$. These different phrasings of the producer's problem are special cases of the different versions of the two-stage formulation of the profit maximization problem developed in Lecture 5, to which we now turn attention.

First, we clarify terminology. A function, $g(x)$, is *linear* in x if

$$g(x) = \sum_{n=1}^{N} a_n x_n,$$

where $(a_1, a_2, ..., a_N)$ are constants. A function, $g(x)$, is *affine* in x if

$$g(x) = c + \sum_{n=1}^{N} a_n x_n, \tag{4}$$

where c and $(a_1, a_2, ..., a_N)$ are constants. Setting $c = 0$ shows that a linear function is a special case of an affine function. When depicted in the (x, y) plane, both a linear function and an affine function have graphs that are straight lines. The graphs of linear functions go through the origin; affine functions can have (positive or negative) intercepts on the vertical axis.

The Profit Function from the Revenue Function

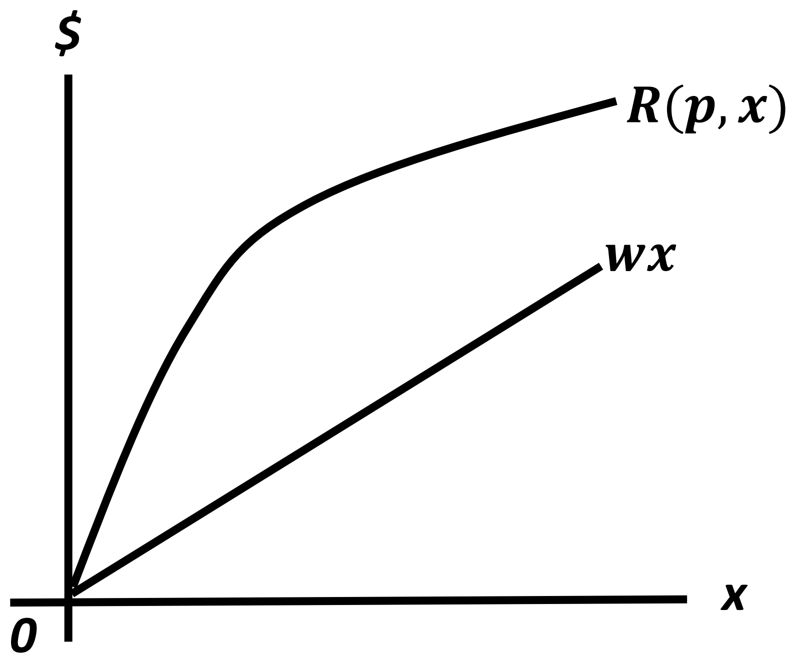
The profit maximization problem framed as

$$\pi(p,w) \equiv \max_x \{R(p,x) - wx\},$$

is illustrated in Figure 5. There the line wx depicts input cost as a linear function of x with a slope of $w > 0$, and $R(p,x)$ depicts maximal revenue as a function of x.[10] Both are

[10]We depict $R(p,x)$ for a canonical technology. That's done for visual familiarity. It is not essential to the argument.

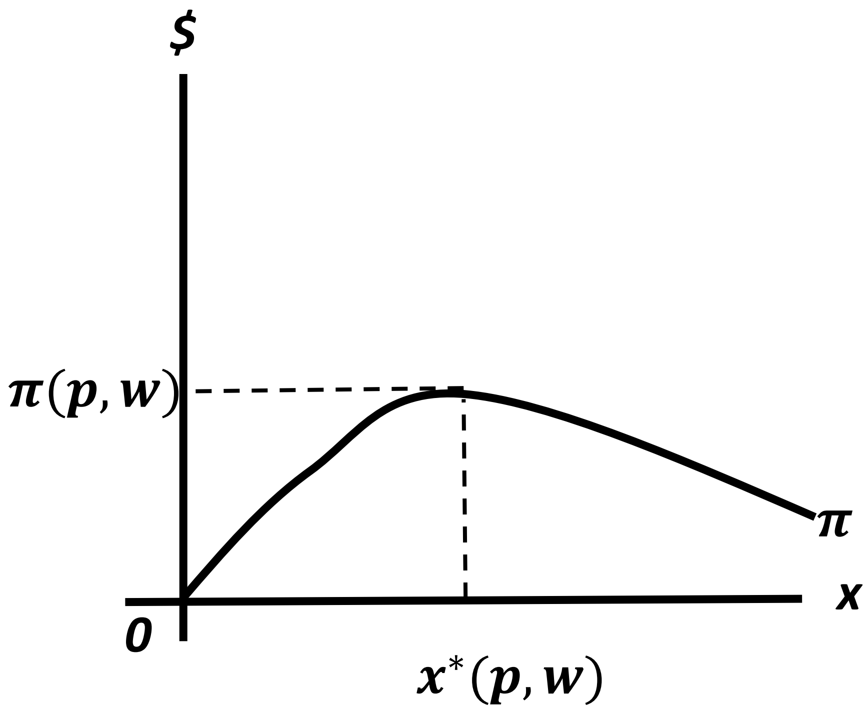
Figure 5: Profit and Revenue



denominated in money units, and profit for each input level is the vertical difference between the revenue curve and the cost line. The problem is to find where that vertical distance is the largest.

We can view the problem from a different perspective by performing the indicated subtraction for each x and then graphing the result in $(x, \$)$ space. Figure 6 illustrates. The

Figure 6: Profit as Function of Input



optimal input demand is at $x^*(p, w)$ that is at the top of the hill-shaped figure. Figures 5 and 6 illustrate a "nice" case where the optimal solution is easy to find. But regardless of how "nice" the case, the principle is the same. One seeks a point where either increasing or decreasing input use reduces profit.

To show how we solve this problem, we note that $x^*(p,w)$ must satisfy

$$R(p,x^*(p,w)) - wx^*(p,w) \geq R(p,x) - wx$$

for all x. In words, if $x^*(p,w)$ is profit maximizing it must generate more profit than any other x. This simply repeats the definition of profit maximizing. Adding wx to both sides of this inequality gives

$$R(p,x^*(p,w)) + w(x - x^*(p,w)) \geq R(p,x) \tag{5}$$

for all x as the condition that the profit-maximizing choice of $x^*(p,w)$ must satisfy.

Expression (5) says that the *affine function* of x,

$$R(p,x^*(p,w)) + w(x - x^*(p,w)) = R(p,x^*(p,w)) - wx^*(p,w) + wx \tag{6}$$

is always greater than or equal $R(p,x)$. (Here the analogue of c in expression (4) is $R(p,x*(p,w)) - wx^*(p,w) \equiv \pi(p,w)$ that does not depend upon x.) Because the graph of an affine function in $(x,\$)$ space is a line (hyperplane), expression (5) requires that line (hyperplane) to be an upper bound for the graph of $R(p,x)$ for all x. That means that the graph of $R(p,x)$ must be contained in the half space below the line (hyperplane). By construction, however, the affine function and $R(p,x)$ share a common value at $x = x^*(p,w)$. Together these facts imply that the affine function must be *tangent to the graph of* $R(p,x)$ *from above at* $x^*(p,w)$.[11] Because the slope (normal) of the affine function is given by w, the tangency condition translates into economic terms as requiring that *the marginal revenue from* x *be equated to* w. Figure 7 illustrates.

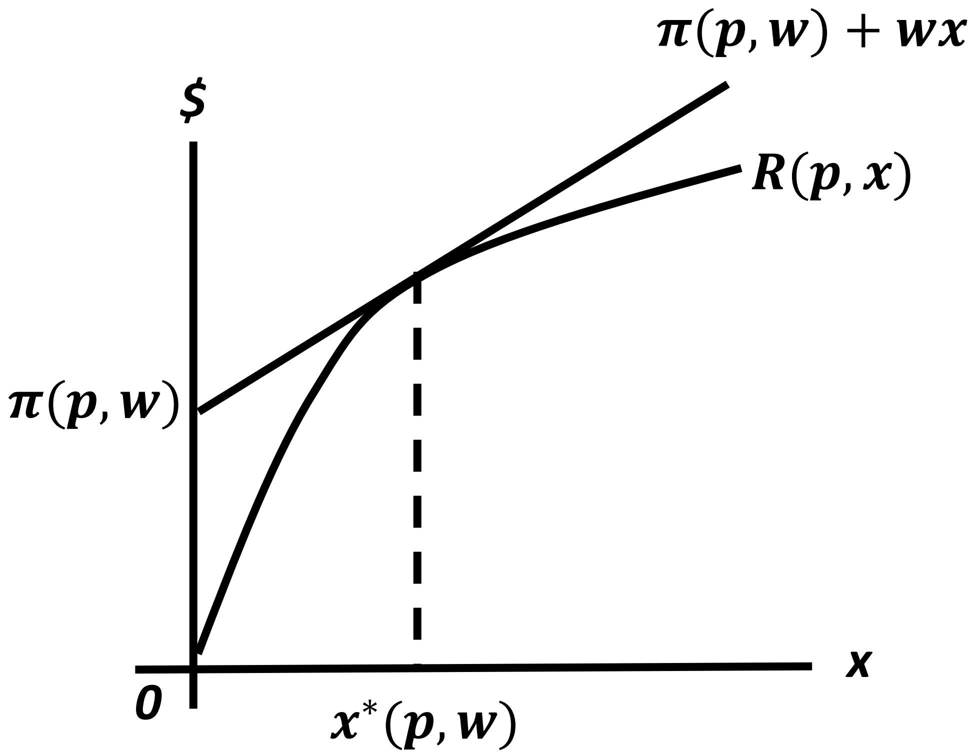
Equating the marginal revenue for an input to its price is the standard economic rule-of-thumb for finding the profit maximizing x. But it's only part of the story. The other part is that the tangency must be *from above* so that the graph of $R(p,x)$ must lie everywhere below that of the affine function in (6). Figure 8 illustrates a situation where the tangency

[11] If we rewrite (5) as

$$w(x - x^*(p,w)) \geq R(p,x) - R(p,x^*(p,w)),$$

and compare it to expression (1) in Lecture 6 (see also the associated footnote), we can recognize this condition as requiring that w be a *supergradient* for $R(p,x)$ at $x^*(p,w)$.

Figure 7: Profit Maximizing Input Demand



condition is satisfied at three points. One is profit maximizing, but two (labelled A and B) are not. (See the dashed line segments with slope w.) Moreover, it's clear from inspection that there are input levels to the right and left of the input associated with A that involve a higher profit than at A. For small enough changes, the same is not true for B.

Point A is an example of what is referred to as a *local minimum* for the difference between revenue and cost. It's precisely what we're trying to avoid. Nevertheless it satisfies the tangency condition. B is an example of a *local maximum.* It's better than points close to it, but it's not the best. And the best is what we're looking for. Condition (5) characterizes it. Equating marginal revenue to w does not.[12]

Mathematicians use the concepts of "necessary" conditions and "sufficient" conditions in these type of discussions. Statement B is called a "necessary" condition for statement A to be true if it can be phrased as follows: If statment A holds then B must be true. Thus, B being a necessary condition for A does not exclude it from holding when A is not true. Instead it only ensures that it must hold when A is true. Statement B is called a "sufficient" condition for A to hold if it can be phrased as follows: If statement B holds then A must be true. In other words, B guarantees that A holds.

So using this terminology, "tangency" is a necessary, but not sufficient, condition because a tangency can occur even if profit is not maximized. A tangency condition is of interest because it always applies at solutions. But because it can apply elsewhere (for example, at local minima), it's not enough.

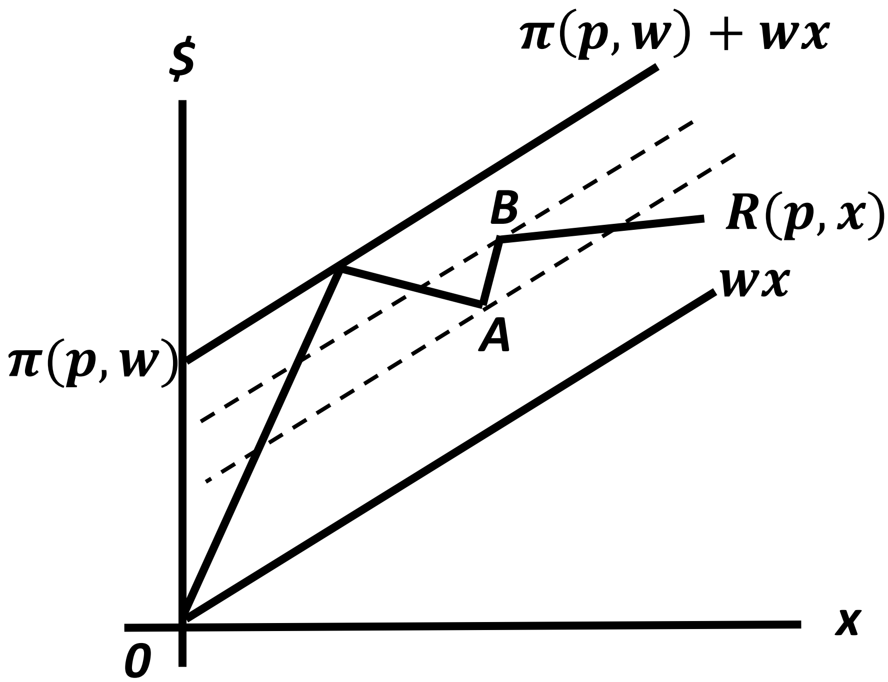
Condition (5), on the other hand, is necessary and sufficient for the proft maximization problem. We show this using the definitions. First, assume that the profit maximization problem has been solved; we have identified $x^*(p,w)$. Then by the definition of a maximum

$$\begin{aligned} R(p, x^*(p,w)) - wx^*(p,w) &= \max_x \{R(p,x) - wx\} \\ &\geq R(p,x) - wx \end{aligned}$$

for all x. This, however, implies that (5) holds verifying that it is necessary for $x^*(p,w)$ to be a maximum. To establish that condition (5) is sufficient, assume that it holds for some

[12]For those with a calculus background, if you smoothed out the kinks, all three satisfy the first order condition for a maximum. B also satisfies the second-order condition, but remains nonoptimal.

Figure 8: Tangency But Not Maximizing



$\hat{x}$, then $R(p,\hat{x}) - w\hat{x} = \pi(p,w)$ by the definition of a maximum, establishing sufficiency.[13] A mathematical short hand for saying that B is necessary and sufficient for A is to say that A occurs *if and only if* B is true, or $A \Leftrightarrow B$.

The solution to this version of profit maximization problem is the same as the solution to the profit maximization problem posed in terms of T. But in this version, we only solve for $x^*(p,w)$. Optimal revenue is obtained by evaluating the revenue function at $x^*(p,w)$, $R(p,x^*(p,w))$. The profit maximizing supplies, $y^*(p,w)$, therefore, equal the revenue maximizing supplies derived in Chapter 7, $y(p,x)$, evaluated at $x = x^*(p,w)$. That is:

$$y^*(p,w) = y(p, x^*(p,w)). \tag{7}$$

Expression (7) opens the door to a deeper analysis of supply response. To illustrate, we examine how supply responds to prices changing from (p^o, w^o) to (p', w'). By (7)

$$y^*(p',w') - y^*(p^o,w^o) = y(p', x^*(p',w')) - y(p^o, x^*(p^o,w^o)).$$

Now add zero in the form of $y(p', x^*(p^o,w^o)) - y(p', x^*(p^o,w^o))$ to get

$$\begin{aligned} y^*(p',w') - y^*(p^o,w^o) &= y(p', x^*(p^o,w^o)) - y(p^o, x^*(p^o,w^o)) \\ &+ y(p', x^*(p',w')) - y(p', x^*(p^o,w^o)) \end{aligned}$$

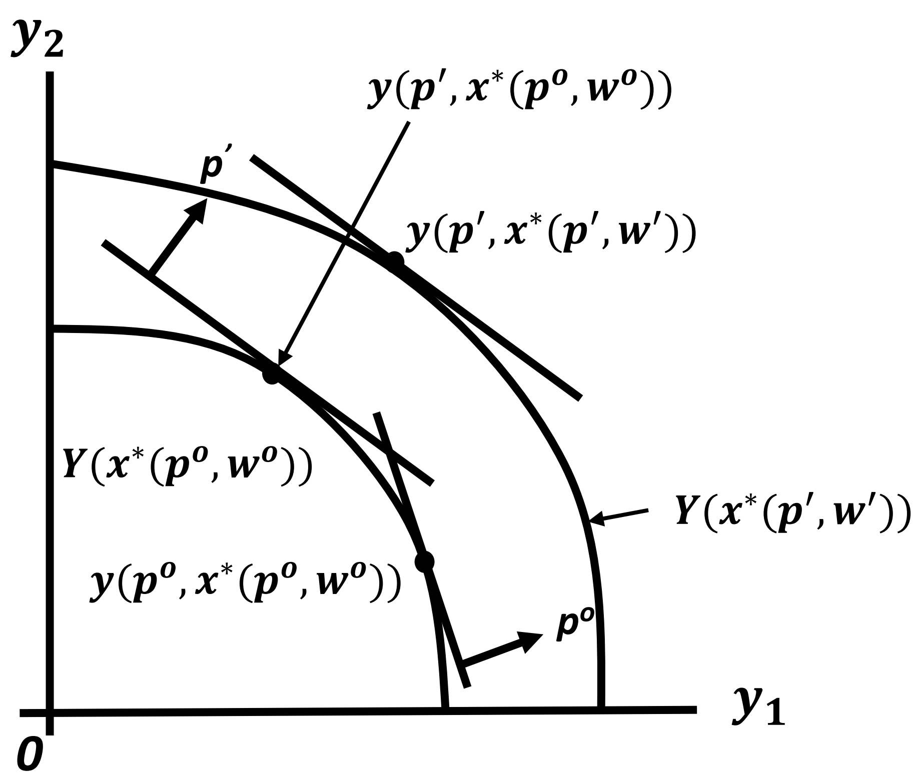
Supply adjustments to changing prices break into two components. One, $y(p', x^*(p^o,w^o)) - y(p^o, x^*(p^o,w^o))$, measures how revenue maximizing supply adjusts to changing output prices from p^o to p' holding input use at $x^*(p^o,w^o)$, the original profit maximizing input demand. It is called the *substitution effect*. Figure 9 illustrates. (Exercise 8 asks you to analyze an alternative way of decomposing this adjustment.)

The second component, $y(p', x^*(p',w')) - y(p', x^*(p^o,w^o))$, measures the supply adjustment caused by the profit maximizing change in input demand. The movement from $y(p', x^*(p^o,w^o))$ on $\bar{Y}(x^*(p^o,w^o))$ to $y(p', x^*(p',w'))$ on $\bar{Y}(x^*(p',w'))$ illustrates. It is called the *expansion effect* because it describes movement across output sets,

The ability to decompose supply adjustments into separate components enhances our ability to parse the causes of observed economic responses. For example, the substitution

[13] As a technical matter, we must verify that a maximum actually exists, but this can always be ensured by making an appropriate assumption.

Figure 9: Output Substitution and Expansion



effect measures revenue maximizing supply response to output price changes. Its behavior is characterized by $R(p,x)$'s properties in p. Economists can predict that behavior. The expansion effect depends on $Y(x)$ and how it adjusts to changes in x. Thus, it is less predictable. The transformation curves in Figure 9 are drawn for a canonical technology, but real-world technologies may not be so orderly.

The Profit Function from the Cost Function

This version of the profit maximization problem is stated

$$\pi(p,w) \equiv \max_{y} \{py - c(w,y)\}.$$

Figure 10 illustrates the problem as finding the output that gives the largest vertical difference between the ray py with slope p, which depicts revenue, and the graph of the cost function, $c(w,y)$.[14] If one does the subtraction in $\{\cdot\}$ and graphs the result in the $(y,\$)$ plane, the result (not drawn) mirrors Figure 6. The difference is that the horizontal axis now measures y and not x.

Because $y^*(p,w)$ is profit maximizing for (p,w), it satisfies

$$py^*(p,w) - c(w, y^*(p,w)) \geq py - c(w,y),$$

for all y. Adding $c(w,y)$ to and subtracting $py^*(p,w) - c(w,y^*(p,w))$ from both sides gives

$$c(w,y) \geq py - [py^*(p,w) - c(w,y^*(p,w))] \tag{8}$$

for all y as the necessary and sufficient condition for profit maximization.

The minimal cost of producing the output, y, must be greater than or equal to the affine function of y,

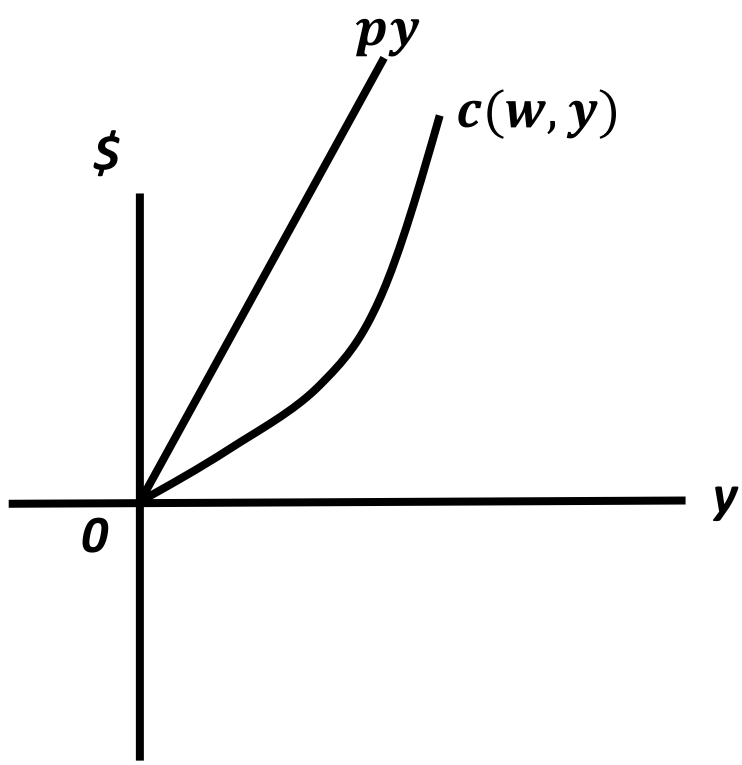
$$py - [py^*(p,w) - c(w,y^*(p,w))], \tag{9}$$

for all y and equal it at $y = y^*(p,w)$. Here the constant of the affine function is

$$-[py^*(p,w) - c(w,y^*(p,w))] = -\pi(p,w).$$

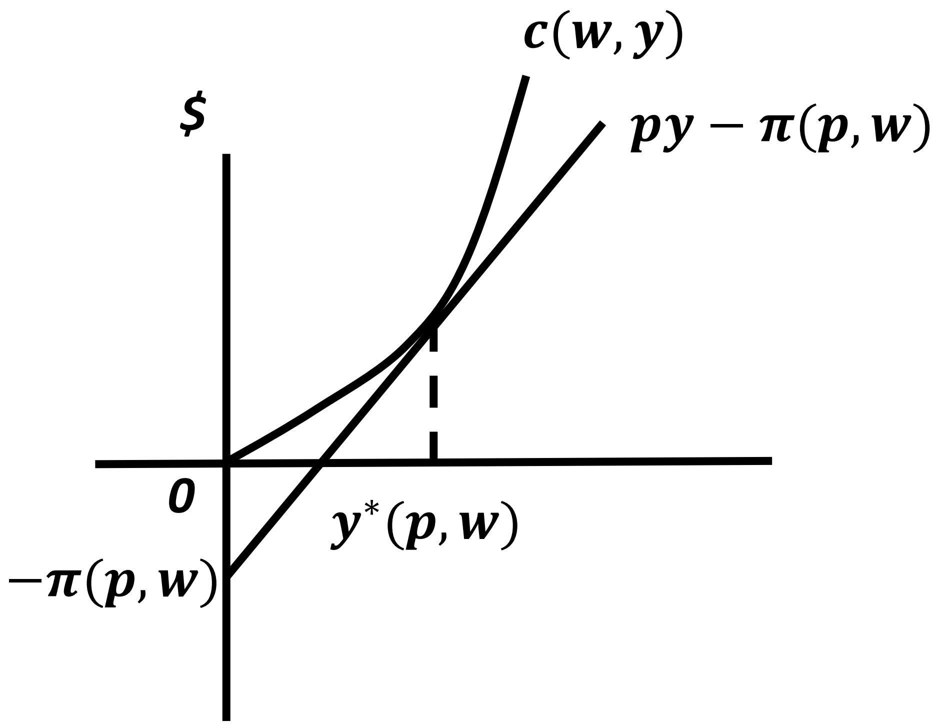
[14] The cost function in Figure 10 is drawn for a canonical technology. This is not essential to the analysis.

Figure 10: Profit Function from Cost Function



Thus, the graph of the cost function in $(y, \$)$ space must be contained in the half space above the graph of this affine function and the two must touch at the profit maximizing y, $y^*(p, w)$. [15] Figure 11 illustrates the tangency for a canonical technology.

Figure 11: Profit Maximizing Supply



The slope of the cost function is *marginal cost*, and the slope (normal) of the affine function in (9) is p. Thus, the familiar economic rule-of-thumb: *Profit maximizing supply is determined by the producer equating price to marginal cost.* While that's true, more is required. This tangency must be from below, and the graph of $c(w, y)$ must lie everywhere above the affine function (9). (Exercise 5 asks you to illustrate a cost structure for which equating price to marginal cost does not ensure a profit maximizing output choice.)

[15] Alternatively, p must be a subgradient of $c(w, y)$ at $(y^*(p, w), c(w, y^*(p, w)))$.

Because producers choose their profit-maximizing inputs to minimize the cost of producing their profit-maximizing output supplies:

$$x^* (p, w) = x (w, y^* (p, w)),$$

profit-maximizing input demands equal cost-minimizing demands evaluated at the profit-maximizing supply. Therefore,

$$x^* (p^o, w^o) - x^* (p, w) = x (w^o, y^* (p^o, w^o)) - x (w, y^* (p, w)).$$

Adding zero in the form of $x (w^o, y^* (p, w)) - x (w^o, y^* (p, w))$ gives

$$\begin{aligned} x^* (p^o, w^o) - x^* (p, w) &= x (w^o, y^* (p, w)) - x (w, y^* (p, w)) \\ &+ x (w^o, y^* (p^o, w^o)) - x (w^o, y^* (p, w)). \end{aligned}$$

Adjustments in input demand, therefore, break into two components. The first, the *substitution effect*,

$$x (w^o, y^* (p, w)) - x (w, y^* (p, w)),$$

captures movement along the isoquant for $y^* (p, w)$ made by a cost minimizing producer in response to input prices changing from w to w^o. The second, the *expansion effect*,

$$x (w^o, y^* (p^o, w^o)) - x (w^o, y^* (p, w)),$$

captures input adjustments across isoquants induced by supply responding to (p, w) change.[16]

[16] Readers familiar with consumer demand theory will recognize that the substitution and expansion effects discussed here have parallels in the substitution and income effects. One difference is that the expansion effect can never result in a profit-maximizing input demand curve sloping upward. There are no Giffen inputs.

Exercises

1. For the single-input, single-output case, the slope of an isoprofit line is determined by the input-output price ratio, $\frac{w}{p}$. For the single-input, single-output case with constant returns (as in Figure 1), draw a family of isoprofit lines for which maximal profit is zero.

2. Figure 4 depicts the "do-nothing" and best-response case for a profit-maximizing producer facing an input-price change. Develop the analogous figure and discussion for an output-price change.

3. Figure 4 illustrates Hotelling's Lemma for input prices. Develop the analogous representation in output-price space.

4. Using Hotelling's Lemma construct a geometric explanation of why convexity of $\pi(p, w)$ implies optimal supplies are upward sloping in their own prices and why input demands are downward sloping in their own prices.

5. For the revenue function depicted in Figure 8 derive the profit-maximizing input demand as w varies over $0 < w < \infty$. Illustrate it in (x, w) space.

6. Illustrate and explain the following decomposition of output response to a price change:

$$\begin{aligned} y^*(p', w') - y^*(p^o, w^o) &= y(p^o, x^*(p', w')) - y(p^o, x^*(p^o, w^o)) \\ &+ y(p', x^*(p', w')) - y(p^o, x^*(p', w')) \end{aligned}$$

7. Illustrate a situation where equating price to marginal cost does not ensure a profit maximizing output choice.

8. Illustrate and explain the following decomposition of input response to an input price change:

$$\begin{aligned} x^*(p^o, w^o) - x^*(p, w) &= x(w, y^*(p^o, w^o)) - x(w, y^*(p, w)) \\ &+ x(w^o, y^*(p^o, w^o)) - x(w, y^*(p^o, w^o)). \end{aligned}$$

Lectures on Neoclassical Production Economics[1]

Lecture 9: Duality

Robert G Chambers[2]

January 9, 2024

[2]University of Maryland

Lecture Introduction

We learned in Lecture 8 that the only technical restriction required to develop the properties of the profit function, $\pi(p, w)$, is that T be nonempty. Despite that, economists routinely model the technology as taking the canonical form. The reason is that *the existence of a positively homogeneous and convex profit function,* $\pi(p, w)$*, implies that a "dual technology",* $\bar{T}$*, that satisfies nonemptiness, free disposability of inputs, free disposability of outputs, and convexity must exist. Moreover,*

$$\pi(p, w) = \max_{x,y} \left\{ py - wx : (x, y) \in \bar{T} \right\}.$$

Furthermore, if the decisionmaker's opportunity cost is zero and prices, (p, w)*, exist for which maximal profit is zero,* $\bar{T}$ *is a canonical technology.* Thus, regardless of the true structure of T, its profit function must be consistent with a canonical technology.

Economists refer to this result as *the profit function,* $\pi(p, w)$*, being "dual" to a canonical technology,* $\bar{T}$. The result's origin is that expressions defined by multiplying prices times quantities define two distinct linear forms. For example, when we write[1]

$$\pi = \sum_{m=1}^{M} p_m y_m - \sum_{n=1}^{N} w_n x_n$$

profit, π, is the dependent variable, but both of the following statements are true:

- A: Profit is a linear function of the independent variables, *quantities* (y, x), with normal given by *prices* $(p, -w)$.

- B: Profit is a linear function of the independent variables, *prices* $(p, w,)$, with normal given by *quantities* $(y, -x)$.

Statement B is formed from A by switching the roles of quantities and prices in statement A. It is is said to be *dual to A*. By the same token, A is *dual* to B. So profit is interpretable as a *linear function of quantities* with prices providing the relevant slopes or as a *linear function of prices* with quantities providing the relevant slopes. What's true for profit is also true for

[1] The expression $\sum_{m=1}^{M} p_m y_m - \sum_{n=1}^{N} w_n x_n$ is an example of what mathematicians also call a *bilinear* form.

cost and revenue. Each is *bilinear* in the sense that is intepretable either as a linear function of quantities or as a linear function of prices.[2]

Duality is a principle in both mathematics and logic. Its exact definition varies. But a common one is that a *duality* translates one true statement into another true statement by exchanging two words or concepts. A and B above illustrate. By convention, economists refer to relationships taking quantities as independent variables *primal* and ones with *prices* as independent variables *dual*.

The "duality" between $\pi(p, w)$ and $\bar{T}$ ensures that, regardless of T's true nature, economists studying price-taking profit maximizers lose no generality by modeling it "as if" it were canonical. This lecture proceeds as follows. We first give an overview of the intuition. Next, we develop the properties of the "dual technology" and examine its relationship to the "true" technology, T. Then we study different manifestations of this duality using optimization principles. The Lecture concludes with some general remarks on dual structures.

The Dual Technology, $\bar{T}$, Defined

Profit maximizing input demands and supplies, $(x^*(p, w), y^*(p, w))$, must be feasible,

$$(x^*(p, w), y^*(p, w)) \in T,$$

and generate more profit than any other feasible input-output bundle,

$$py^*(p, w) - wx^*(p, w) \geq py - wx$$

for all $(x, y) \in T$ and all prices (w, p). Substituting the definition of the profit function, $\pi(p, w)$, into the left-hand side of the inequality gives

$$\pi(p, w) \geq py - wx \tag{1}$$

for all $(x, y) \in T$ as a restriction on feasible input-output bundles.

[2]*The two-dimensional real coordinate plane and its N-dimensional extension are special cases of what mathematicians refer to as a *vector space*. For an arbitrary vector space, X, its *dual space*, denoted X^*, is the space of linear functions over X. For both the two-dimensional real coordinate plane and its N-dimensional extension, its dual space is the space itself.

In the single-output case, expression (1) can be rewritten as

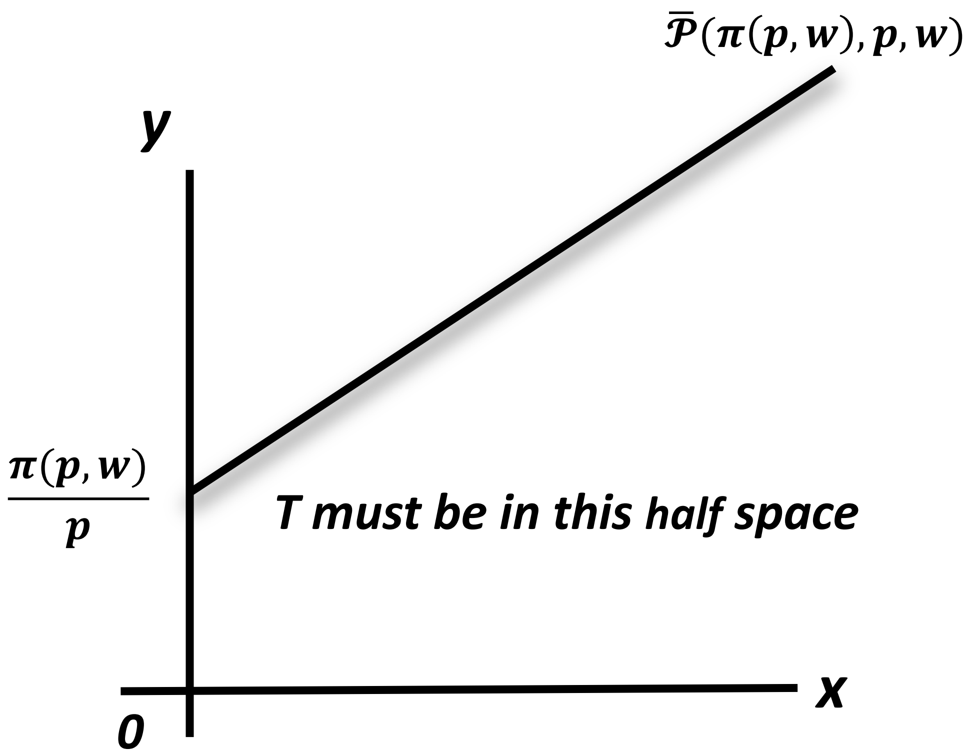
$$y \leq \frac{\pi(p,w)}{p} + \frac{w}{p}x$$

for all $(x, y) \in T$. This inequality requires that any (x, y) belonging to T must lie on or below the isoprofit line (hyperplane) defined by maximal profit, $\pi(p, w)$,

$$\bar{\mathcal{P}}(\pi(p,w), p, w) = \{(x, y) : \pi(p,w) = py - wx\},$$

that Figure 1 illustrates as the line segment with vertical intercept $\frac{\pi(p,w)}{p}$. Because $\pi(p, w)$ represents the maximal profit available from T given (p, w), T cannot contain any input-output pairs lying above the line segment. Expression (1), however, holds for *all possible*

Figure 1: Isoprofit Line for $\pi(p, w)$



input and output prices. Therefore, if we keep p constant (make it the numeraire), and vary the input price over, say, $w^2 > w^1 > w^0$, the same logic implies that T must fall below each of the resulting isoprofit lines. Because T must lie below each pf these isoprofit lines, it must fall in the *intersection* of the half spaces below them. That is,

$$T \subset \mathcal{P}^- \left(\pi \left(p, w^o\right), p, w^o\right) \cap \mathcal{P}^- \left(\pi \left(p, w^1\right), p, w^1\right) \cap \mathcal{P}^- \left(\pi \left(p, w^2\right), p, w^2\right).$$

Figure 2 depicts this intersection as the (x, y) pairs lying on or below the piecewise linear contour $\frac{\pi\left(p,w^2\right)}{p}ABC$.

Figure 2: Constructing $\bar{T}$

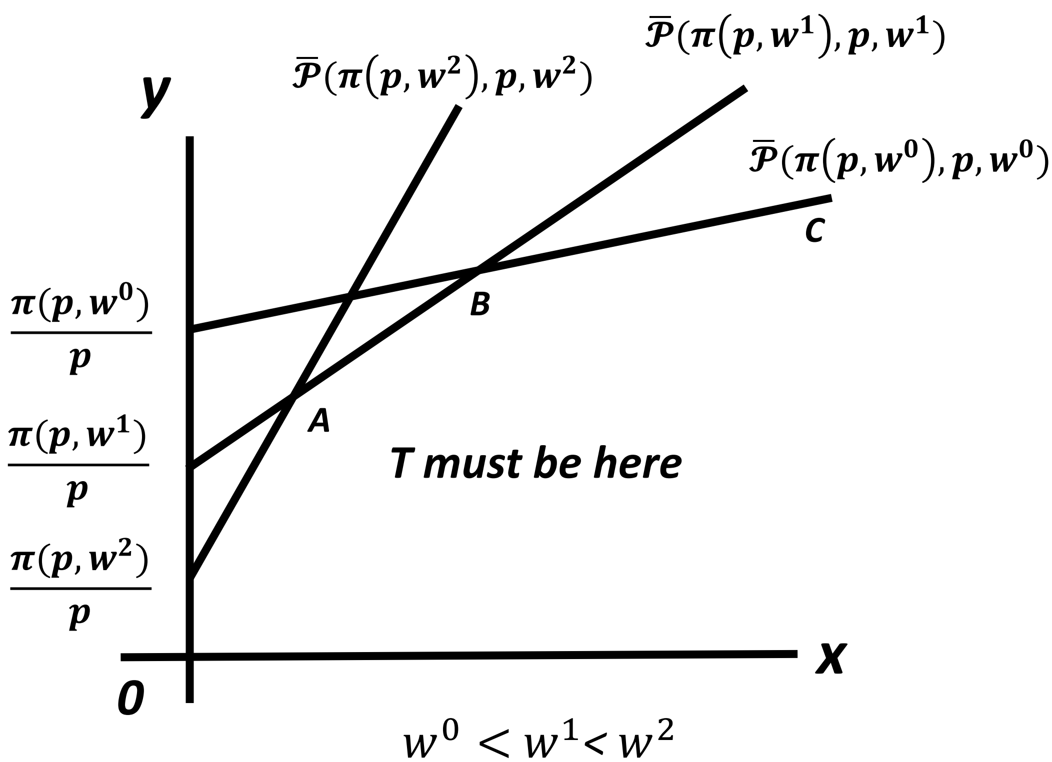


Figure 2 is drawn for three input price levels. Letting w vary between $(0, \infty)$, however, partitions off all the half spaces in which T must fall. (Because only real prices matter, letting w vary over that range is the same as letting the real price $\frac{w}{p}$, which determines the slope of isoprofit lines, range over $(0, \infty)$). These arguments suggest the following theorem. (The theorem is true regardless of the dimension of x or y):

Theorem 1. *(Fundamental Duality Theorem)* $T \subset \bar{T}$ where

$$\begin{aligned} \bar{T} &\equiv \{(x, y) : py \leq \pi(p, w) + wx \text{ for all } (p, w)\} \\ &= \cap_{p,w} \{(x, y) : py \leq \pi(p, w) + wx\}. \end{aligned}$$

Theorem 1 identifies the set of inputs and outputs, $\bar{T}$, that we call the "dual technology". We prove it using a proof by contradiction: *Suppose that, contrary to Theorem 1, a technically feasible input-output bundle,* $(x, y) \in T$*, exists that does not belong to the dual technology,* $(x, y) \notin \bar{T}$*. Because it does not fall in* $\bar{T}$*, at least one* (p, w) *must exist for which* $py - wx > \pi(p, w)$*. This inequality requires that a technicaly feasible* (x, y) *exists that yields a higher profit than maximal available profit,* $\pi(p, w)$*, which gives the desired contradiction. Hence, Theorem 1 must be true.*

The intuition is that knowledge of the profit function can be used to "reverse-engineer" the profit maximization problem to isolate a region of input-output bundles in which the technology must lie. The search for $\pi(p, w)$ assumes that the producer "knows" T and wants to find the highest possible isoprofit line (hyperplane) touching it. On the other hand, in constructing $\bar{T}$, we use knowledge of $\pi(p, w)$ and (p, w) to find the slopes and intercepts of the isoprofit lines (hyperplanes) under which T must fall.

Properties of the Dual Technology, $\bar{T}$

The area below the piece-wise linear $\frac{\pi(p, w^2)}{p} ABC$ in Figure 2 is nonempty, satisfies free disposability of inputs and outputs (FDI and FDO), and is a convex set. We first show that these properties hold in general. Then we discuss conditions that ensure $\bar{T}$ satisfies NFL and NFC.

To verify that $\bar{T}$ satisfies FDI, we must show that

$$x' \geq x \text{ for } (x, y) \in \bar{T} \Rightarrow (x', y) \in \bar{T}.$$

Let $(x, y) \in \bar{T}$ so that $py - wx \leq \pi(p, w)$ for all (p, w). Taking $x' \geq x$ shows that

$$py - wx' \leq py - wx \leq \pi(p, w) \text{ for all } (p, w),$$

which ensures that $(x', y) \in \bar{T}$. For $\bar{T}$ to satisfy FDO, it must be true that if $(x, y) \in \bar{T}$ then $(x, y') \in \bar{T}$ for $y' \leq y$. Let $y' \leq y$ for $(x, y) \in \bar{T}$ so that $py - wx \leq \pi(p, w)$ for all (p, w). Then we must have

$$py' - wx \leq py - wx \leq \pi(p, w) \text{ for all } (p, w),$$

which ensures $(x, y') \in \bar{T}$.[3]

To show that $\bar{T}$ is convex, we must show that if $(x^o, y^o) \in \bar{T}$ and $(x^*, y^*) \in \bar{T}$, then any weighted average of (x^o, y^o) and (x^*, y^*) also belongs to $\bar{T}$. That is, we need to prove that

$$\begin{array}{rcl} & (x^o, y^o) \in \bar{T} \quad \text{and} \quad (x^*, y^*) \in \bar{T} & \\ & \Downarrow & \\ p(\lambda y^o + (1-\lambda) y^*) - w(\lambda x^o + (1-\lambda) x^o) & \leq & \pi(p, w) \text{ for all } (p, w), \end{array}$$

for $\lambda \in (0, 1)$. Because $(x^o, y^o) \in \bar{T}$ and $(x^*, y^*) \in \bar{T}$,

$$\begin{aligned} py^o - wx^o &\leq \pi(p, w) \text{ for all } (p, w) \\ py^* - wx^* &\leq \pi(p, w) \text{ for all } (p, w) \end{aligned}$$

Multiply both sides of the first inequality by $\lambda > 0$, the second by $(1 - \lambda) > 0$, and add the results to get

$$p(\lambda y^o + (1-\lambda) y^*) - w(\lambda x^o + (1-\lambda) x^o) \leq \pi(p, w) \text{ for all } (p, w),$$

which is what we needed to show.

That leaves no free lunch (NFL) and no fixed costs (NFC). Here, we assume that the decisionmaker's opportunity cost is 0 and that a set of prices exist for which $\pi(p, w) = 0$. If NFL is violated there exists an $(0, y) \in T$ with $y \geq 0$ and $y \neq 0$. That input-output bundle generates a profit of $py - w0 > 0$, which cannot satisfy

$$py - wx \leq \pi(p, w) \ \texttt{for all} \ (p, w)$$

if a (p, w) exists such that $\pi(p, w) = 0$.[4] Hence, $\bar{T}$ is consistent with NFL.

[3] For both FDI and FDO recall that input and output prices are assumed positive.

[4] Again recall that all output prices are strictly positive.

If the decisionmaker's opportunity cost is zero, the decisionmaker will not produce at prices for which $\pi(p,w) < 0$. Thus, observed profit for those prices is zero but never negative, and

$$0 = p0 - w0 \leq \pi(p,w) \texttt{ for all } (p,w),$$

so that $(0,0) \in \bar{T}$. $\bar{T}$ is thus consistent with NFC.

What is the Dual Technology $\bar{T}$?

If the decisionmaker's opportunity cost is 0, and a set of prices exist for which $\pi(p,w) = 0$, the dual technology, $\bar{T}$, is canonical. We know, however, that a well-behaved profit function, $\pi(p,w)$, exists even if the technology, T, looks like Figure 4 in Lecture 5. Therefore, $\bar{T}$, in general, is not the same set as T. Nonetheless,

$$\bar{\pi}(p,w) \equiv \max_{x,y} \left\{ py - wx : (x,y) \in \bar{T} \right\}$$

satisfies

$$\bar{\pi}(p,w) = \pi(p,w)$$

for all (p,w). Although $\bar{T}$ and T can differ, they have the same profit function.

To help explain, Figure 3 reproduces Figure 3 from Lecture 8. Recall that T there consists of all (x,y) on or below the curve AY. It does not satisfy no fixed costs and, therefore, is not canonical. For the illustrated prices, maximal profit, $\pi(p,w)$, for T is negative, which is illustrated by its isoprofit line having a negative intercept. A decisionmaker facing this T and a zero opportunity cost foregoes production at these prices in favor of alternative employment (see Lecture 8). Because the decisionmaker does not act as a producer, the input-output bundle at these prices is zero and not $(x^*(p,w), y^*(p,w))$. By analogy, any real price $\frac{w}{p}$ for an isoprofit line that supports T from above but that has a negative intercept on the vertical axis leads the decisionmaker to forego production. Positive production is weakly preferred to the opportunity cost when the real price of the input falls to the level $\frac{\bar{w}}{p}$ depicted in Figure 4 as the slope of $\bar{\mathcal{P}}(0,p,\bar{w})$. For that real input price, the profit maximizing input-output bundle in T, labelled as point $\bar{A}$, produces a zero profit. The decisionmaker is indifferent between being a producer and alternative employment.

Figure 3: No Fixed Costs Violated

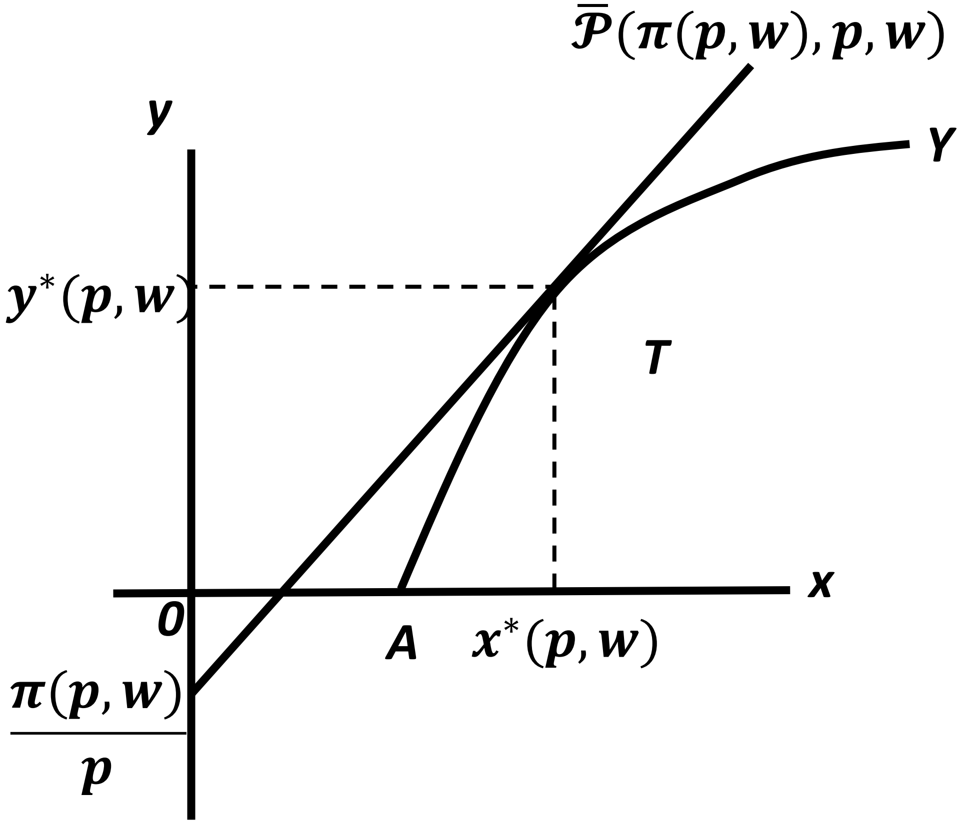
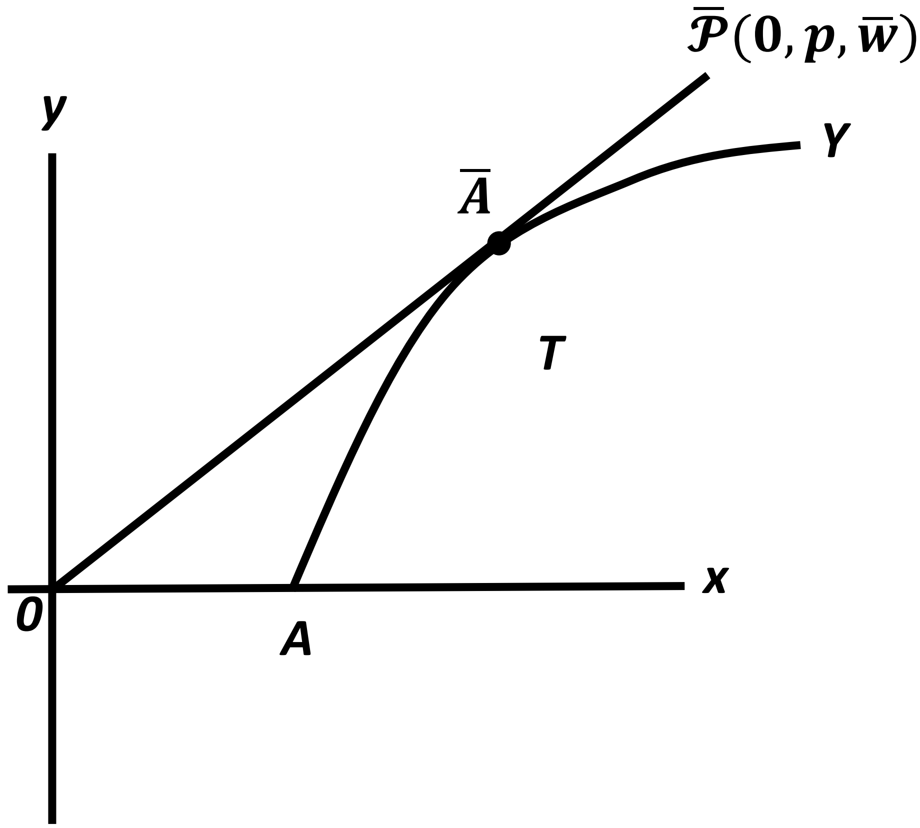


Figure 4: Observed Profit



Real input prices lower than $\frac{\bar{w}}{p}$, whose isoprofit lines are "flatter" than $\bar{\mathcal{P}}(0, p, \bar{w})$, will support positive production of output with a positive profit. As the real price of the input falls, the producer moves along the boundary of T from $\bar{A}$ towards Y. Therefore, if the technology is the area on or below the curve AY in Figure 4, a decisionmaker with zero opportunity cost never produces along $A\bar{A}$ because that involves negative profit. The decisionmaker is better off not being a producer for those prices.

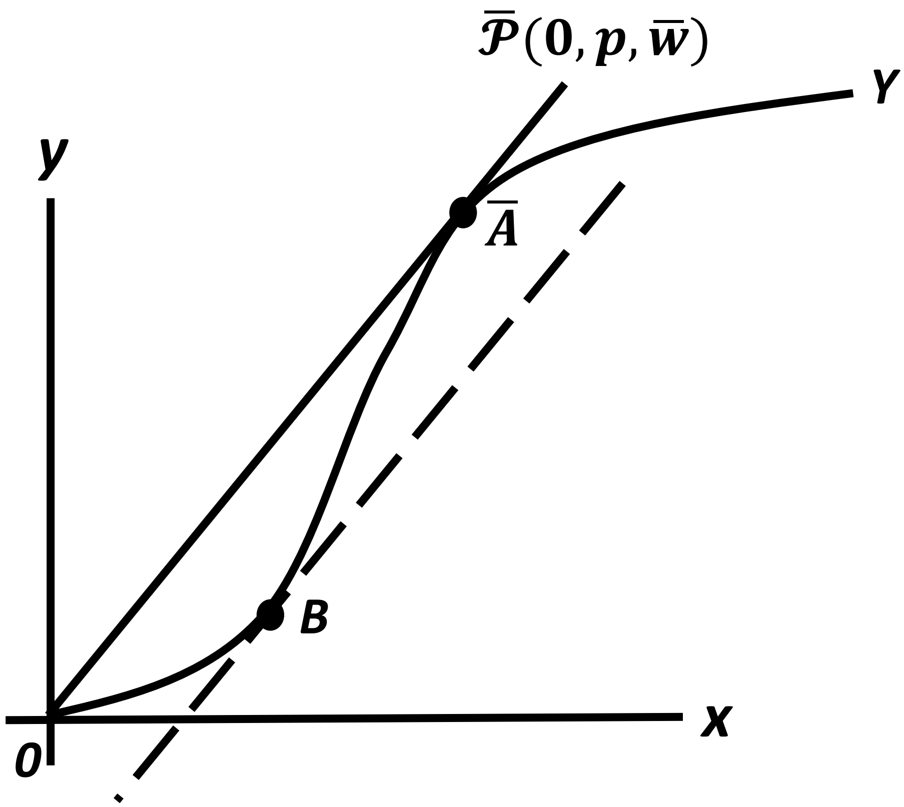
Now let's compare that behavior with that of a decisionmaker who faces a technology given by the area on or below $0\bar{A}Y$ in Figure 4. (That is, the graph of the production function for this technology would be the line segment $0\bar{A}$ and then $\bar{A}Y$ afterwards.) For real prices of the input higher than $\frac{\bar{w}}{p}$, whose isoprofit lines have "steeper" slope than $\bar{\mathcal{P}}(0, p, \bar{w})$, profit maximization occurs at $(0, 0)$. For real input prices lower than $\frac{\bar{w}}{p}$, profit maximization occurs along the arc $\bar{A}Y$. The production pattern is the same as that for a profit maximizer who faces T in Figure 3 and has an opportunity cost of 0. The area on or below $0\bar{A}Y$ is the $\bar{T}$ that would be constructed if the technology were that in Figure 3. It contains T as a subset and gives the same profit function.

Figure 5 offers another perspective. There T is everything on or below the curve $0Y$. This is the variable returns technology introduced in Figure 7 of Lecture 3. For real input prices higher than $\frac{\bar{w}}{p}$, which is depicted as the slope of the isoprofit line $\bar{\mathcal{P}}(0, p, \bar{w})$, the profit maximizing solution is $(0, 0)$.

If, however, you relied on the rule-of-thumb of finding a tangency between the isoprofit line and the boundary of T to determine the profit maximizing point, you might think otherwise. Isoprofit lines can be tangent from above to the boundary of T for real prices higher than $\frac{\bar{w}}{p}$ in the region $B\bar{A}$. However, T is not fully contained in the half spaces below such isoprofit lines. Thus, these isoprofit lines fall below the one passing through $(0, 0)$. Tangencies can also can occur along $0B$ for real input prices lower than $\frac{\bar{w}}{p}$. But these will occur from below. Drawing their isoprofit lines shows that higher isoprofit lines can be attained in the region $\bar{A}Y$, while staying in T.

If a decisionmaker faces the technology in Figure 5, profit maximizing production occurs either at $(0, 0)$ or along its upper frontier in the region $\bar{A}Y$. If you compare the profit levels for a decisionmaker facing that technology with one facing a technology given by everything

Figure 5: Nonconvex Technology



on or below $0\bar{A}Y$, you will find that they are same. In this case, $\bar{T}$ corresponds to everything below $0\bar{A}Y$. Again $T \subset \bar{T}$ and gives the same profit function.

The phenomena captured in Figures 3 through 5 is general. $\bar{T}$ does not give the technology. Instead it gives the smallest possible set in input-output space that satisfies: nonemptiness, FDI, FDO, convexity, NFC, NFL *and that contains the true* T. Moreover, it has the same maximal profit as T for all prices (p, w). Figures 4 and 5 illustrate why. Profit maximizing decisionmakers who have an opportunity cost of zero never operate in regions of T that are not consistent with a canonical technology. It's not that such regions never exist. To the contrary, we have recognized since Lecture 2 that they can. But profit maximizing decisionmakers avoid them in favor of regions that are consistent with a canonical technology.[5]

The importance of this result is hard to overestimate. It carries immense analytic advantages. First, it confirms that the competitive profit-maximizing choices for $\bar{T}$ will correspond to those for T. Therefore, in studying the behavior of competitive profit maximizing producers, we get the same answers, whether we study $\bar{T}$ or T. Thus, we can analyze producer behavior "as if" they respond to a canonical technology. The two representations are *observationally equivalent* for competitive profit maximizers. Second, it ensures that no generality is lost in studying profit maximizing producer behavior by first specifying a well-behaved profit function. If the $\pi(p, w)$ is appropriately specified, there must exist a corresponding canonical $\bar{T}$ that encapsulates all the relevant economic information about the true technology, T, regardless of the underlying properties of T.

Other Manifestations of Duality

Because $\bar{T}$ satisfies satisfies the properties of a canonical technology, we can use it to develop both a cost function and a revenue function. That cost function would be nondecreasing and convex in outputs, and that revenue function would be nondecreasing and concave in inputs.

[5] $\bar{T}$ is called the *free-disposal convex hull of* T meaning that it is the smallest set consistent with the properties of a canonical technology that contains T as a subset.

Instead of pursuing this approach, we investigate a cost function that is derived from the profit function as

$$\bar{c}(w,y) \equiv \max_{p} \{py - \pi(p,w)\} \tag{2}$$

and a revenue function derived from the profit function as

$$\bar{R}(p,x) \equiv \min_{w} \{wx + \pi(p,w)\} \tag{3}$$

that we refer to, respectively, as the *dual cost function* and the *dual revenue function.* Our focus in the dual cost function discussion is on it as a function of outputs, y, and the focus in the dual revenue discussion is on it as a function of inputs, x. Exercises 5 and 6 at the end of this lecture ask you to show that $\bar{c}(w,y)$ is positively homogeneous and concave as a function of w and that $\bar{R}(p,x)$ is positively homogeneous and convex as a function of p. Thus, the "dual" structures satisfy the same properties in prices as the ones derived directly from T.

The Dual Cost Function, $\bar{c}(w,y)$

Solving the Dual Cost Problem

The definition of the profit function, $\pi(p,w)$, requires that

$$\pi(p,w) \geq py - c(w,y) \tag{4}$$

for all (p,y,w). A tiny bit of algebra applied to (4) gives

$$c(w,y) \geq py - \pi(p,w),$$

for all (p,y,w) as an immediate consequence. The difference between revenue from y and maximal profit, $\pi(p,w)$, is always less than or equal to the minimal cost of producing y. That is, *the difference between revenue from y and maximal profit, $\pi(p,w)$ provides a lower bound for the minimal cost of producing y.* This holds for all (p,y,w). And, in particular, it holds *when keeping y and w fixed, we choose p to maximize the difference between revenue from y and $\pi(p,w)$.* Thus,

$$\begin{aligned} c(w,y) &\geq \max_{p} \{py - \pi(p,w)\} \\ &\equiv \bar{c}(w,y) \end{aligned}$$

for all (w, y). The dual cost function, $\bar{c}(w, y)$, is a lower bound to $c(w, y)$. On the other hand, if y equals $y^*(p^o, w)$,

$$c(w, y^*(p^o, w)) = p^o y^*(p^o, w) - \pi(p^o, w).$$

But that means that p^o must solve the version of (2) framed as

$$\bar{c}(w, y^*(p^o, w)) = \max_p \{py^*(p^o, w) - \pi(p, w)\},$$

which is the dual cost problem for the profit maximizing supply $y^*(p^o, w)$.

Therefore, the dual cost function, $\bar{c}(w, y)$, is a lower bound for the cost function, $c(w, y)$, that equals it whenever y is profit maximizing for (p, w). That leaves open the possibility that $\bar{c}(w, y) > c(w, y)$ when y is not a profit maximizing vector of output supplies. This happens because $\bar{c}(w, y)$ is the minimal cost function for $\bar{T}$, not T. (We don't show this, but it is true.) The divergence between $\bar{c}(w, y)$ and $c(w, y)$ reflects the divergence between $\bar{T}$ and T.

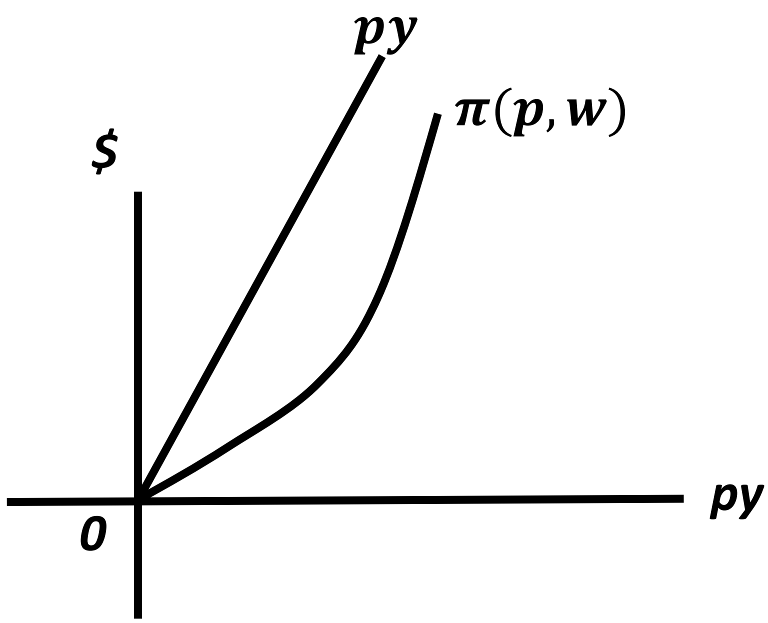
Figure 6 illustrates the dual cost problem (2). We now seek a p that makes the vertical difference between the ray from the origin labelled py and the graph of the profit function $\pi(p, w)$ as large as possible. Note the similarity to Figure 10 of Lecture 8. The problem structures are the same even though the problems posed differ. The profit maximization problem is posed in (primal) quantity space and maximizes profit given (dual) prices. The dual cost problem is posed in (dual) price space and maximizes the difference between revenue for given (primal) output and $\pi(p, w)$. The expression py is interpreted differently in the two problems. In Figure 10 of Lecture 8, it is a linear function of y with normal p. Those roles are reversed in Figure 6. But despite the *different intuitive interpretations*, the inherent mathematical structures of both problems are identical. Therefore, we use the same math techniques to solve problem (2) that we used to solve the profit maximization problem. The difference is that we now search over p instead of over y.

Denote price vectors that solve the dual cost minimization problem by

$$p(w, y) \in \underset{p}{\operatorname{argmax}} \{py - \pi(p, w)\}.$$

Here, $\operatorname{argmax}_p \{\cdot\}$ is an instruction to choose any price vector that gives the largest possible

Figure 6: Deriving $\bar{c}(w, y)$



value in the set $\{\cdot\}$ and, thus, solves problem (2). Any solution[6] to the dual cost problem satisfies

$$py - \pi(p, w) \leq p(w, y) y - \pi(p(w, y), w) \tag{5}$$

for all p with equality for $p = p(w, y)$. Thus, for $p(w, y)$ to solve the dual cost problem, y must be the normal for the affine function of p,

$$\pi(p(w, y), w) + y(p - p(w, y)) = py - \bar{c}(w, y),$$

whose graph supports the graph of $\pi(p, w)$ from below at $(p(w, y), \pi(p(w, y)))$. Figure 7 illustrates.[7]

The definition of the dual cost function, $\bar{c}(w, y)$, also requires that

$$\bar{c}(w, y^*) \geq p(w, y) y^* - \pi(p(w, y), w)$$

for all y^* and y and

$$\bar{c}(w, y) = p(w, y) y - \pi(p(w, y), w).$$

Subtracting the inequality from the inequality gives

$$\bar{c}(w, y^*) \geq \bar{c}(w, y) + p(w, y)(y^* - y) \tag{6}$$

for all y^*. Any solution to the dual cost minimization problem, $p(w, y)$, must be a normal for an affine function of y^* whose graph supports the graph the dual cost function from below at $(y, \bar{c}(w, y))$.[8]

We summarize expressions (5), (6), and the surrounding discussions in statement C below.

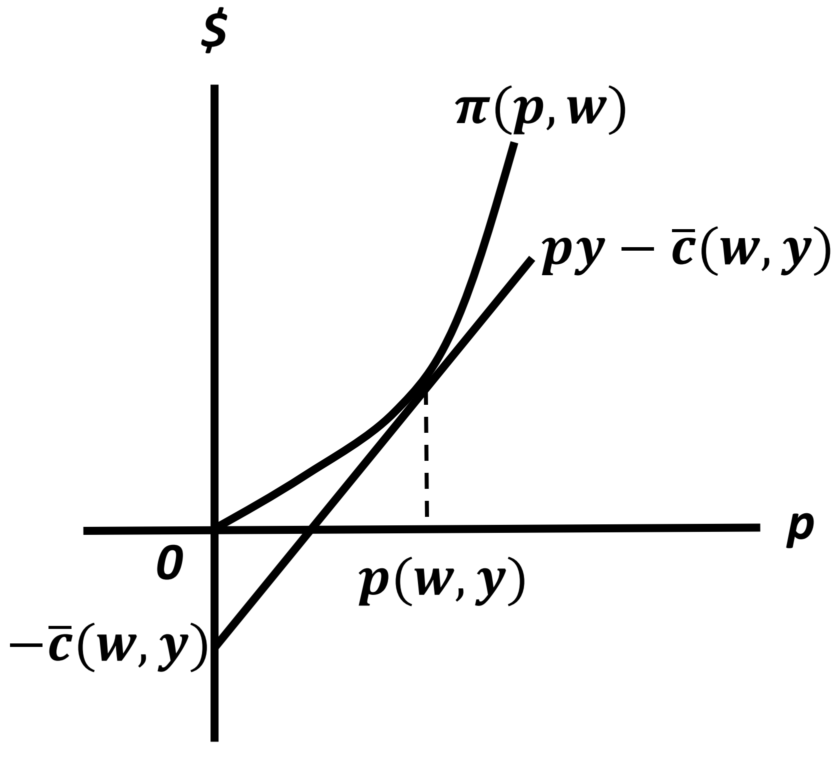
- C: The solution to the dual cost problem occurs where y forms the normal for a hyperplane that supports the profit function from below at $(p, \pi(p, w))$, and any solution to the dual cost problem forms the normal for a hyperplane that supports the graph of the dual cost function from below at $(y, \bar{c}(w, y))$.

[6] We say *any* solution because the problem can have multiple solutions. If that occurs, all the solutions give the same value of $py - \pi(p, w)$.

[7] $p(w, y)$, thus, solves the dual cost problem if and only if y is a *subgradient* of $\pi(p, w)$ in p.

[8] $p(w, y)$ is a *subgradient* of $\bar{c}(w, y)$.

Figure 7: Solution to Dual Cost Problem (Hotelling's Lemma)



Switching the roles of "dual cost" and "profit" and y and p in C gives its dual

- D: The solution to the profit problem occurs where p forms the normal for a hyperplane that supports the dual cost function from below at $(y, \bar{c}(w, y))$, and any solution to the profit problem forms the normal for a hyperplane that supports the graph of the profit function from below at $(p, \pi(p, w))$.

The first (second) part of D (C) restates the condition developed in Lecture 8 (price equals marginal cost) that must be satisfied for y to solve

$$max_y \{py - \bar{c}(w, y)\},$$

and the second (first) part of D (C) restates part of Hotelling's Lemma. Both problem solutions give the same information.

Dual Cost as a Function of Output

We now show that $\bar{c}(w, y)$ is nondecreasing (nonnegative marginal cost) and convex in output (nondecreasing marginal cost), y. Because we obtain the dual cost function by solving a problem whose form is identical to that of the profit maximization problem but involves different choice variables (p now instead of y), our method for demonstrating these properties mirrors that used to establish that $\pi(p, w)$ is nondecreasing and convex in p.

We first show that $y' \geq y$ implies

$$\bar{c}(w, y') \geq \bar{c}(w, y).$$

To do so, we again deploy a "do nothing" argument. If $y' \geq y$ it must be true that

$$p(w, y)\, y' - \pi(p(w, y), w) \geq p(w, y)\, y - \pi(p(w, y), w).$$

The left hand side is the result of "doing nothing" to the solution of the dual cost problem in response to output increasing from y to y'. Because output prices, $p(w, y)$, are positive, increasing output while holding them constant forces revenue to increase. Because the dual cost problem is to maximize the difference between revenue and maximal profit, the best response must dominate the "do nothing" response and cannot increase dual cost.

For convexity, let

$$\hat{y} = \lambda y^0 + (1-\lambda) y^1,$$

for $\lambda \in (0,1)$ denote a weighted average of two output bundles y^0 and y^1. To demonstrate that the dual cost structure is convex in output, we must show that the weighted average of $\bar{c}(w, y^0)$ and $\bar{c}(w, y^1)$ always is at least as large as the dual cost of $\hat{y}$, $\bar{c}(w, \hat{y})$. By the definition of the dual cost function:

$$\begin{aligned} \bar{c}\left(w, y^0\right) &\geq p(w, \hat{y}) y^0 - \pi(p(w, \hat{y}), w) \\ \bar{c}\left(w, y^1\right) &\geq p(w, \hat{y}) y^1 - \pi(p(w, \hat{y}), w) \end{aligned}$$

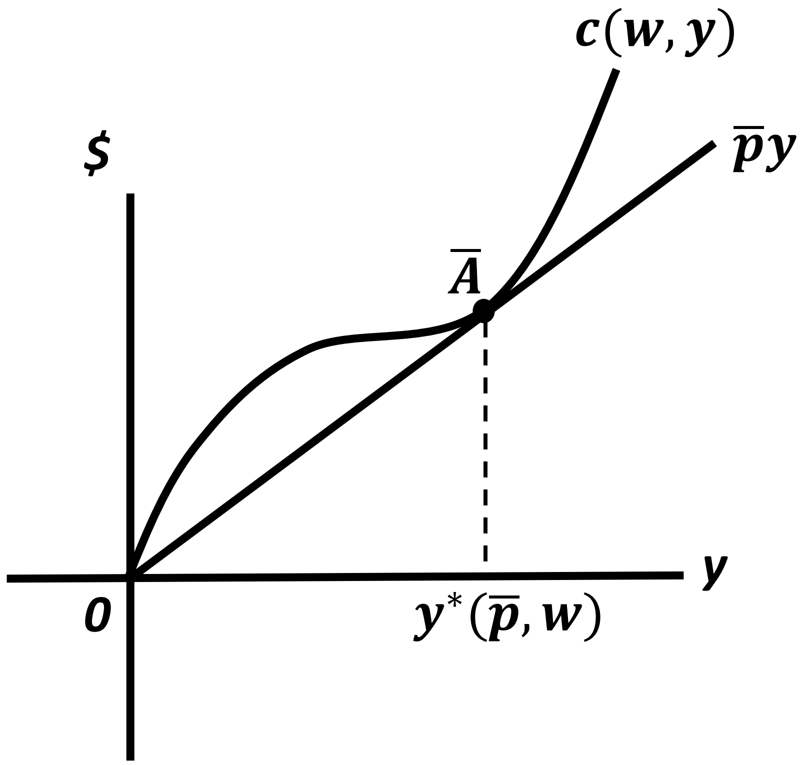
Multiply the first inequality by $\lambda > 0$, the second by $(1-\lambda) > 0$, and add the result to obtain

$$\begin{aligned} \lambda\bar{c}\left(w, y^0\right) + (1-\lambda)\bar{c}\left(w, y^1\right) &\geq p(w, \hat{y})\left[\lambda y^0 + (1-\lambda) y^1\right] - \pi(p(w, \hat{y}), w) \\ &= \bar{c}(w, \hat{y}) \end{aligned}$$

for $\lambda \in (0,1)$.

Figure 8 illustrates why $\bar{c}(w, y)$ and $c(w, y)$ do not coincide. As drawn, the illustrated cost function initially has decreasing marginal costs and then increasing marginal costs. If the price of the output is below $\bar{p}$, the graph of the ray from the origin that depicts revenue will always lie below the graph of the cost function. That means that profit is negative for all output levels except zero. The producer is best off, therefore, operating at the origin, producing nothing, and realizing a cost and profit of zero. For price $\bar{p}$, the producer makes zero profit by producing $y^*(\bar{p}, w)$ and is, therefore, indifferent between producing that amount and zero. For prices higher than $\bar{p}$, price can equal marginal cost in the region $0\bar{A}$ of the graph of the cost function. But such points fall on lower isoprofit lines than the parallel tangencies that occur in the region beyond $\bar{A}$. Thus, a profit maximizer never produces in the region $0\bar{A}$ of the cost function. And because $\bar{c}(w, y)$ is constructed from profit maximizing solutions via $\pi(p, w)$, the closest approximation to $c(w, y)$ in that region is the shut-down solution that would occur if the slope of the cost curve coincided with $\bar{p}$ between 0 and $y^*(\bar{p}, w)$. The graph of $\bar{c}(w, y)$ in $(y, \$)$ space thus follows the line segment $0\bar{A}$ until $y^*(\bar{p}, w)$

Figure 8: Nonconvex Cost and Profit Maximization



and the graph of $c(w,y)$ afterwards. And as required, $c(w,y) \geq \bar{c}(w,y)$ and both always involve the same profit.

Dual Revenue from the Profit Function

Solving the Dual Revenue Problem

By the definition of the profit and revenue functions,

$$\pi(p,w) \geq R(p,x) - wx$$

for all (p,x,w). Some easy algebra gives

$$\pi(p,w) + wx \geq R(p,x)$$

for all (p,x,w). Because this inequality must hold for all possible values of $\pi(p,w) + wx$, it must hold when w makes it as small as possible. Therefore,

$$\min_{w} \{\pi(p,w) + wx\} \geq R(p,x),$$

for all (p,x). Hence,

$$\bar{R}(p,x) \geq R(p,x),$$

for all (p,x) and

$$\bar{R}(p,x^*(p,w)) = R(p,x^*(p,w)).$$

The dual revenue function is an upper bound for the revenue function for T, and the two revenue functions coincide for profit maximizing input vectors.

We use the same strategy to solve and analyze the dual revenue problem. As a result, the discussion is more terse. Denote input-price vectors that solve the dual revenue problem by:

$$w(p,x) \in \operatorname*{argmin}_{w} \{\pi(p,w) + wx\}.$$

Here $\text{argmin}_w \{\cdot\}$ is an instruction to choose the vector of input prices that gives the smallest possible value in the set $\{\cdot\}$. For given x, any such solution (there can be more than one) must satisfy:

$$\pi(p,w(p,x)) - x[w - w(p,x)] \leq \pi(p,w), \tag{7}$$

for all w. The interpretation of this condition is, by now, familiar. For an input-price vector $w(p,x)$ to solve the dual revenue problem, the graph of the affine function of w,

$$\pi(p, w(p,x)) + w(p,x)x - xw,$$

with normal $-x$ must support the graph of $\pi(p,w)$ from below at $(w(p,x), \pi(p, w(p,x)))$.[9] Expression (7) is illustrated by Figure 4 from Lecture 8, by taking x^* and w^* there to be, respectively, x and $w(p,x)$. Similarly, taking $p^* = p$, $w^* = w(p,x)$, and $x^*(p^*, w^*) = x$ in expression (1) in Lecture 8 gives expression (7), the second half of Hotelling's Lemma.

By the definition of the dual revenue function:

$$\begin{aligned} \bar{R}(p,x^*) &\leq \pi(p, w(p,x)) + w(p,x)x^*, \ \textit{and} \\ \bar{R}(p,x) &= \pi(p, w(p,x)) + w(p,x)x, \end{aligned}$$

for all x^* and x. Hence,

$$\bar{R}(p,x^*) \leq \bar{R}(p,x) + w(p,x)(x^* - x) \tag{8}$$

for all x^*. Expression (8) requires that the graph of the affine function of x^*

$$\bar{R}(p,x) + w(p,x)(x^* - x)$$

with normal $w(p,x)$ must support the graph of $\bar{R}(p,x^*)$ from below at $\left(x, \bar{R}(p,x)\right)$.

We summarize (7), expression (8), and the surrounding discussion in statement D.

- E: The solution to the dual revenue problem occurs where $-x$ forms the normal for a hyperplane that supports the graph of the profit function from below at $(w, \pi(p,w))$, and any solution to the dual revenue problem forms the normal for a hyperplane that supports the graph of the dual revenue from below at $\left(x, \bar{R}(p,x)\right)$.

Switching the roles of "dual revenue" and "profit" and $-x$ and w in C gives its dual

- F: The solution to the profit problem occurs where w forms the normal for a hyperplane that supports the graph of the dual revenue function from below at $\left(x, \bar{R}(p,x)\right)$, and (minus) any solution to the profit problem forms the normal for a hyperplane that supports the graph of the profit function from below at $(w, \pi(p,w))$

[9] In words, $-x$ must be a *subgradient* of $\pi(p,w)$ in w at the solution.

The first (second) part of F (E) restates the condition that must be satisfied for an output vector to be profit maximizing (price equals marginal cost), and the second (first) part of F (E) restates part of Hotelling's Lemma.

Dual Revenue as a Function of Input

It remains to show that the dual revenue function, $\bar{R}(p,x)$, is nondecreasing and concave in the input vector x. The method used is analogous to that used to show that the cost function is nondecreasing and concave in w. That $\bar{R}(p,x)$ is nondecreasing in x follows from another "do nothing" argument. Suppose that in response to the input bundle falling from, say, x' to x, nothing is done to the solution of the dual revenue problem for x'. Then

$$\pi(p, w(p, x')) + w(p, x')\, x' \geq \pi(p, w(p, x')) + w(p, x')\, x$$

for $x' \geq x$. Because the best response to the change in x must be no worse than the "do nothing" response, $\bar{R}(p, x') \geq \bar{R}(p, x)$ as required.

To demonstrate concavity, let $\hat{x} = \lambda x^0 + (1-\lambda)\, x^1$ for $\lambda \in (0,1)$. By the definition of the dual revenue function, we must have

$$\begin{aligned} \bar{R}\left(p, x^0\right) &\leq w(p, \hat{x})\, x^0 + \pi(p, w(p, \hat{x}))\,, \textit{ and} \\ \bar{R}\left(p, x^1\right) &\leq w(p, \hat{x})\, x^1 + \pi(p, w(p, \hat{x},))\,. \end{aligned}$$

Multiply the first inequality by $\lambda > 0$, the second by $(1-\lambda) > 0$, and add the results to obtain

$$\begin{aligned} \lambda \bar{R}\left(p, x^0\right) + (1-\lambda)\, \bar{R}\left(p, x^1\right) &\leq w(p, \hat{x}) \left(\lambda x^0 + (1-\lambda)\, x^1\right) + \pi(p, w(p, \hat{x})) \\ &= \bar{R}(p, \hat{x}) \end{aligned}$$

for $\lambda \in (0,1)$ establishing the desired concavity. If you set $p = 1$ for the single output case, the curve $0Y$ in Figure 5 depicts the revenue function for a nonconvex technology, $R(p,x)$. In that case, $\bar{R}(p,x)$ corresponds to $0\bar{A}Y$.

A Road Not Taken?

A reasonable reaction to these duality results might be: *If one knows that self-interested, price-taking producers behave as though there exists a canonical technology, why not start with $\bar{T}$ instead of T?*

One reason is that economists originally conceptualized producers as maximizing the difference between $pf(x)$ and wx, and then reasoned forward. It was gradually established that the conditions required for this problem to have well-behaved solutions implied that a producer's supply curves slope upward in the output price and that input demands slope downward in their prices. Until *circa* 1970, most production analyses took that perspective.

A byproduct of the Second World War was an increased interest in the mathematics of optimization. In the 1950s and 1960s, that brought new perspectives on and new tools for studying maximization and minimization problems. The realization of the importance of the connections between *dual optimization problems* of the type exemplified by

$$\max_{y} \{py - c(w, y)\}$$

and

$$\max_{p} \{py - \pi(p, w)\}$$

was an important outgrowth. So, for many years, economists just did not understand that many economic problems could be studied from a dual perspective.

So, history is a partial answer. But more important is the fundamental idea that model formulation best proceeds by first understanding the basics of the practical problem and then formulating a tractable analytic model. For many economic problems, that means starting from a primal perspective. As important as mathematics is to economic analysis, its ultimate usefulness stems from the *extramathematical reasoning*[10] that the economist attaches to the abstract mathematical objects.

For example, we've learnt that economic behavior consistent with the laws of supply and demand implies the existence of a mathematical object that we've named $\bar{T}$ and interpreted as a canonical technology. $\bar{T}$'s existence, however, is a *purely mathematical consequence* of

[10] A phrase that I borrowed from a famous decision theorist, the late Peter Fishburn.

the existence of a positively homogeneous and convex function of variables (p, w) that we've called "prices". If those objects were, say, different colored marbles, the mathematical object $\bar{T}$ would still exist. But its extramathematical interpretation differs. It assumes its peculiar meaning to us because we framed our model in terms of a set T interpreted as a technology. In their own way, the duality results confirm Le Tellier's observation: *...plus nous avançons dans la connaissance de l'univers, plus il nous apparaît fondè sur des lois mathèmatiques.*[11]

[11] For a translation, see Lecture 1.

Exercises

1.* For an arbitrary $w \in \mathbb{R}^N_{++}$ define the half space:

$$\left\{x \in \mathbb{R}^N : wx \geq c(w, y)\right\}.$$

Draw a two-dimensional representation of that half space. Now take another $w' \in \mathbb{R}^N_{++}$, define the half space it generates as

$$\left\{x \in \mathbb{R}^N : w'x \geq c(w', y)\right\},$$

and illustrate it visually on the same figure. Prove that their intersection,

$$\left\{x \in \mathbb{R}^N : wx \geq c(w, y)\right\} \cap \left\{x \in \mathbb{R}^N : w'x \geq c(w', y)\right\},$$

forms a convex set consistent with free disposability of input that contains $X(y)$ as a subset. Extend this reasoning to show that

$$\begin{aligned} X(y) &\subset \{x : wx \geq c(w, y) \quad \text{for all } w\} \\ &\equiv \bar{X}(y), \end{aligned}$$

the *dual input set*.

2.* Prove that $\bar{X}(y)$ defined in Exercise 1 satisfies free disposability of inputs (FDI) and is a convex set.

3.* Following the reasoning in Exercise 1 to show that

$$\begin{aligned} Y(x) &\subset \{y : py \leq R(p, x) \quad \text{for all p}\} \\ &\equiv \bar{Y}(x), \end{aligned}$$

the *dual output set*.

4.* Prove that $\bar{Y}(x)$ satisfies free disposability of output and is a convex set.

5*. Prove that $\bar{c}(w, y)$ is positively homogeneous and concave as a function of w.

6*. Prove that $\bar{R}(p, x)$ is positively homogeneous and convex as a function of p.