

Econ 802

Lecture Notes on Chapter 1

Greg Dow

Sept. 17, 2021

The first several chapters in Varian are on The Theory of The firm

Our basic assumptions will be:

- ① There are competitive markets for all inputs and outputs (The firm is a price-taker)
- ② The firm chooses quantities of inputs and outputs subject to a technology constraint
- ③ The firm maximizes profit.

Our tasks will be:

- ① derive the implications of profit max for any reasonable technology (what happens to firm behavior as the prices of inputs and outputs change?)
- ② Investigate what happens with specific technologies like Cobb-Douglas or Leontief
- ③ Eventually (in the last part of the course) we will look at the aggregate behavior of many firms interacting in markets. We will endogenize the prices of inputs and outputs using equilibrium conditions.

Chapter 1 is focused on how to describe technology.

A production plan is a vector $y = (y_1, \dots, y_j, \dots, y_n)$
where there are goods $j = 1 \dots n$.

If $y_j > 0$ Then j is a net output of the firm

If $y_j < 0$ Then j is a net input to the firm.

All we care about is the net quantity leaving the firm (for an output) or the net quantity used up (for an input).

Example: suppose $y_1 =$ kg of flour per hour
 $y_2 =$ loaves of bread baked
per hour

If 1 kg of flour can be used to produce 2 loaves of bread and there are no other goods Then
 $y = (-1, 2)$ is a feasible production plan.

Now suppose labor is needed to bake bread.

Let $y_3 =$ person-hours of labor used.

If it takes 3 people working together for an hour to turn 1 kg of flour into 2 loaves of bread Then
 $y = (-1, 2, -3)$ is a feasible production plan.

Note that there is always a time period involved (it may be implicit) so you should think about the quantities of inputs and outputs as flows per unit of time.

Why use the notation of production plans?

- ① it is convenient when there are multiple outputs (don't need a production function)
- ② we can use set theory to describe technology constraints
- ③ we have a nice vector representation of profit. Write $p = (p_1, \dots, p_n) > 0$ as the vector of prices. Then

$$p \cdot y = \sum_{j=1}^n p_j y_j = \text{total revenue from outputs minus total expenditure on inputs}$$

So $p \cdot y = \text{profit}$.

(because inputs are negative, they are subtracted off)

Production possibilities set.

Suppose there are n goods and a typical production plan is $y = (y_1, \dots, y_n)$.

The set of all feasible production plans is Y .

This is called the production possibilities set.

So y is feasible if and only if $y \in Y$.

Note that I will sometimes abbreviate "if and only if" as "iff"

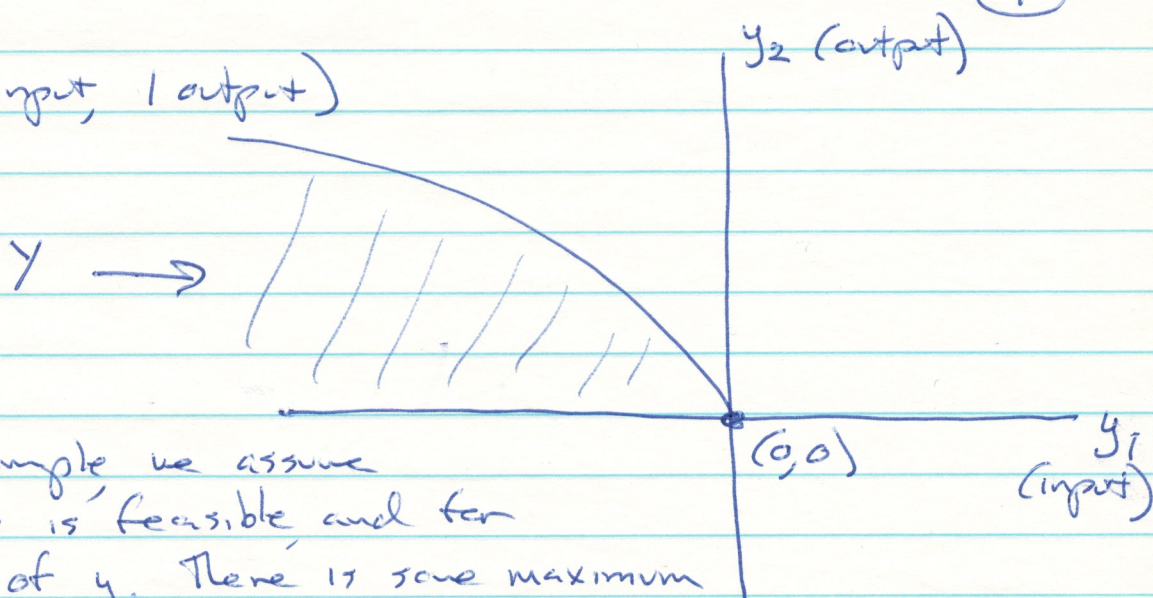
Y is a subset of $\mathbb{R}^n$ the set of all n -dimensional vectors whose entries are real numbers.

The firm chooses the $y \in Y$ that gives the highest possible value of $p \cdot y$. So it solves the problem

$$\boxed{\max_{y \in Y} p \cdot y}$$

(4)

Graph (1 input, 1 output)



In this example we assume any $y_1 \leq 0$ is feasible and for each value of y_1 there is some maximum possible level of y_2 . Smaller outputs $y_2 \geq 0$ are also feasible.

We will normally assume that free disposal of outputs is feasible. So if it would be possible to produce some amount of an output, it is also possible to produce any smaller amount.

In most situations, however, free disposal of outputs only applies to non-negative quantities (like $y_2 \geq 0$). Often it makes no physical sense to have a negative amount.

Long run versus short run

In the long run, all elements of y are variable and any $y \in Y$ can be chosen.

In the short run, one or more elements of y may be fixed.

For example if $y_1 = \text{capital} = -10$, and this is fixed in the short run, then only vectors of the form $y = (-10, y_2, \dots, y_n) \in Y$ are feasible.

These vectors form a subset of Y which we could call $Y(-10)$.

In general, if some sub-vector of inputs is fixed at levels z we could write the short run production possibilities set as $Y(z)$ to indicate that it is conditional on z .

The Input Requirement Set

Assume we have 1 output and n inputs.

Now we change notation and write a production plan as $(y, -x)$ where $y \geq 0$ is a scalar output level and $x = (x_1, \dots, x_n) \geq 0$ is a vector of n input levels.

The input requirement set is defined as

$$V(y) = \{x \in \mathbb{R}_+^n : (y, -x) \in Y\}$$

So the (non-negative) input vector x is in the set $V(y)$ iff there is some feasible production plan such that the output y can be produced using the inputs x .

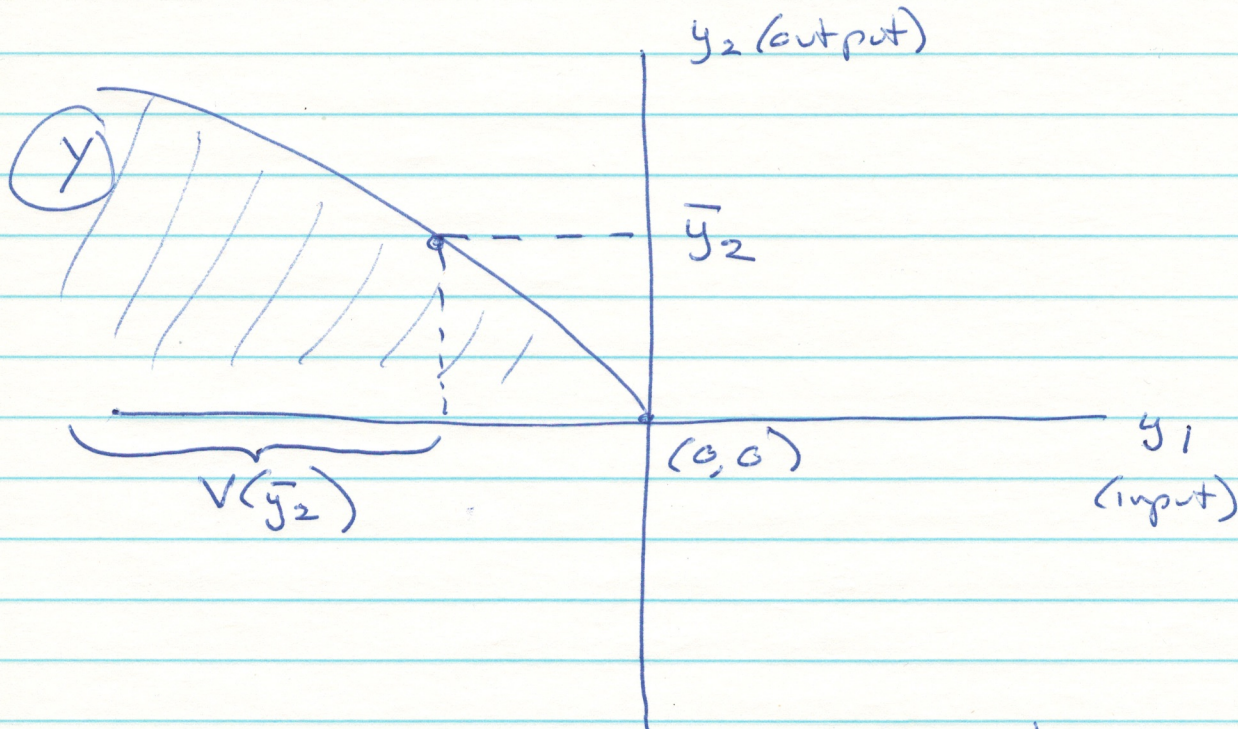
Or to put it more simply: $V(y)$ is the set of all input vectors that could be used to produce y .

Note: $\mathbb{R}_+^n$ is the set of all non-negative n -dimensional vectors.

Also note the change in dimensionality: here Y has dimension $n+1$ and $V(y)$ has dimension n .

(6)

Graphical example: Consider the same graph we used earlier for the 2-dimensional set Y .



Assuming that extra output can always be thrown away and the firm wants the output level $\bar{y}_2$, it can achieve this using any level for y_1 that is in the part of the horizontal axis indicated by $V(\bar{y}_2)$.

Note that Y is 2-dimensional and $V(\bar{y}_2)$ is one-dimensional. Also note that due to the sign convention for inputs in $V(y)$, we use the absolute value of y_1 .

I'll show a case where $V(y)$ is 2-dimensional after I define the n -isoguant.

An isoguant $Q(y)$ is defined as follows:

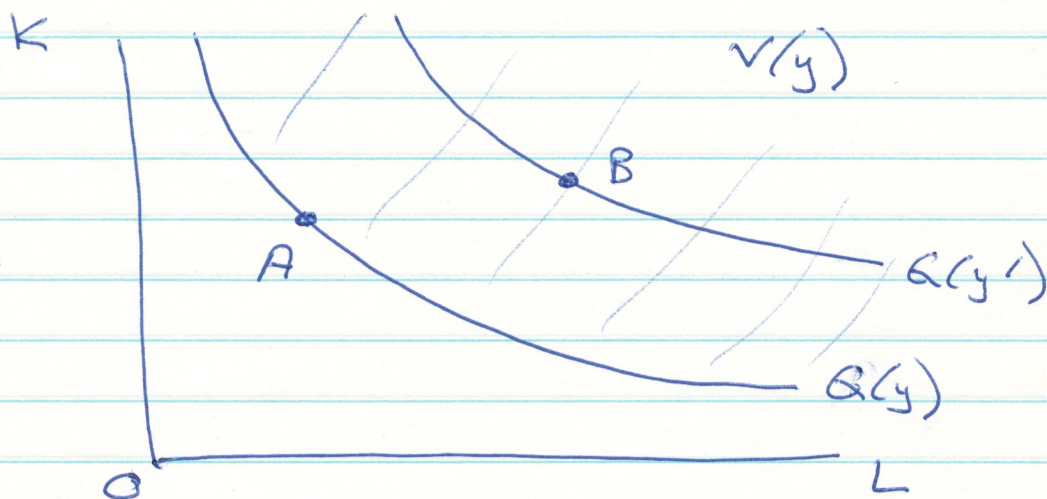
$$Q(y) = \{x \in \mathbb{R}_+^n : x \in V(y) \text{ but } x \notin V(y') \text{ for all } y' > y\}$$

(7)

In words, This says That x is an element of the isoguant set for y if it can be used to produce y , but it cannot be used to produce any output greater than y .

Graph:

Suppose we have two inputs, labor ($L \geq 0$) and capital ($K \geq 0$).



Suppose we want to produce the output level $y > 0$.

You are probably familiar with the standard isoquant curve $Q(y)$ that shows all the (L, K) combinations that yield exactly y .

In a case like this it would make sense for $V(y)$ to include both the curve $Q(y)$ and all the points above it, because points above $Q(y)$ can be used to produce some $y' > y$ and then the extra output can be thrown away. This is an example of a 2-dimensional input requirement set.

For these reasons, points like A and B are both in $V(y)$. However, $A \in Q(y)$ while $B \notin Q(y)$. Point B is not an element of $Q(y)$ because it can be used to produce the larger output $y' > y$. Therefore $B \in V(y')$. According to the definition, this means that B is excluded from $Q(y)$. Of course, we do have $B \in Q(y')$.

8

One more point on this: $Q(y)$ is always a subset of $V(y)$. That is $Q(y) \subseteq V(y)$.

This follows from the definition: in order to have $x \in Q(y)$ we must first have $x \in V(y)$, and then some additional conditions must hold.

The Production Function

Continue to assume there is a single scalar output $y \geq 0$ and there are n non-negative inputs $x = (x_1, \dots, x_n) \geq 0$.

We define the production function as follows:

$$f(x) = \{y \text{ in } \mathbb{R} : y \text{ is the maximum output associated with } -x \text{ in } Y\}$$

To put it another way: for a given x , we look at all vectors in Y of the form $(y, -x)$. Choose the one with the maximum value of y , and assign this value of y to the function $f(x)$.

For examples, see the Cobb-Douglas and Leontief cases on p. 4 of Varian.

Features of a "Reasonable" Technology

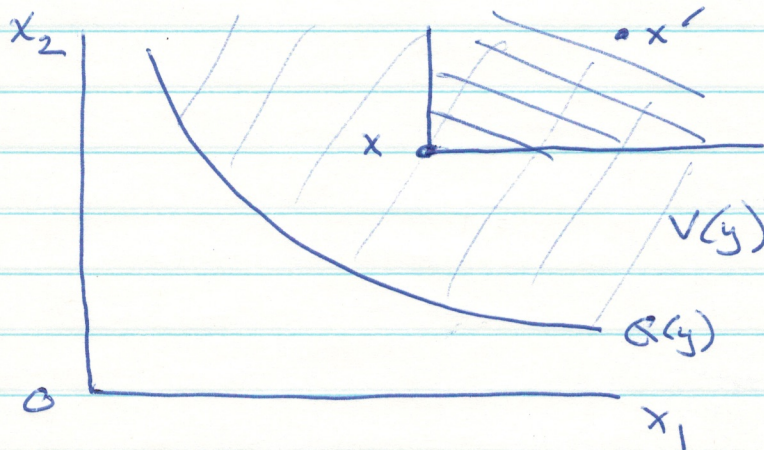
Continue to assume a single output; we are interested in the properties of the input requirement sets $V(y)$.

What would be some reasonable features that most input requirement sets could be assumed to have?

(9)

① Monotonicity (also called free disposal of inputs)

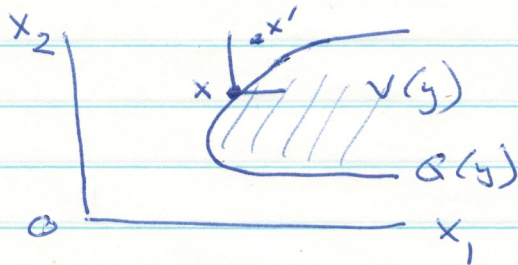
If $x \in V(y)$ and $x' \geq x$, Then $x' \in V(y)$



Let $V(y)$ be the lightly shaded area, where x is a point in $V(y)$. Monotonicity says that every x' to the northeast of x (with at least as much of every input) is also in $V(y)$.

Notice that when monotonicity holds, the set $V(y)$ is not bounded above.

What this assumption rules out is something like this:



where some points to the northeast of x are outside $V(y)$. This doesn't make sense because if you start with x' , you can throw away some inputs to get x , and we have $x \in V(y)$.

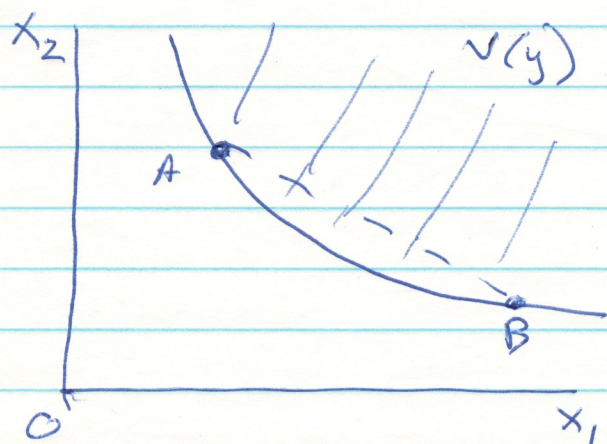
② Convexity: $V(y)$ is a convex set for every $y \geq 0$.

Technical definition: For any $x \in V(y)$ and $x' \in V(y)$ we have $tx + (1-t)x' \in V(y)$ for all $0 \leq t \leq 1$.

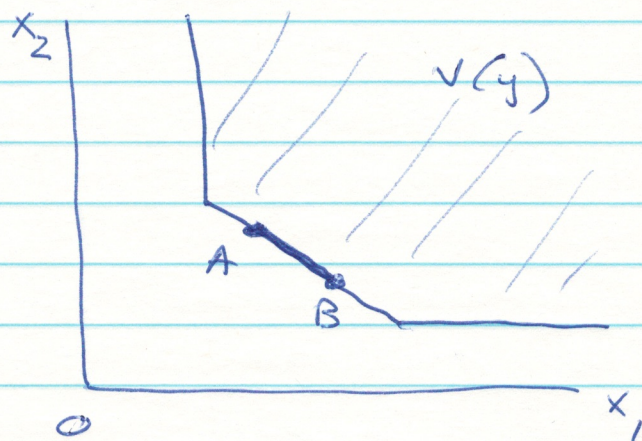
In words: any point along the line segment between x and x' must also be in $V(y)$.

We say that the point $tx + (1-t)x'$ is a convex combination of x and x' . Every point on the line segment from x to x' can be written in this form for some t such that $0 \leq t \leq 1$.

Graphs:

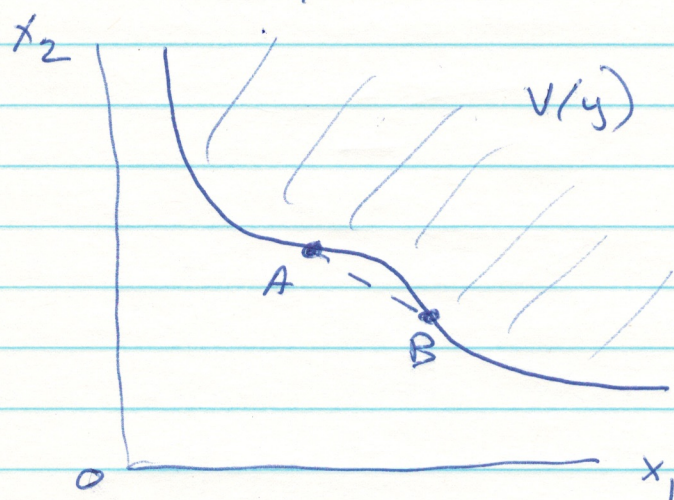


Strictly convex



convex, not strictly

↓
Note that points on the line segment between A and B are on the boundary of $V(y)$, not in the interior.



not convex

For $V(y)$ to be strictly convex, we need $tx + (1-t)x'$ to be in the interior of $V(y)$ for all $0 < t < 1$.

Facts about convexity:

① If Y is convex then $V(y)$ is also convex
(There is a simple proof in Varian)

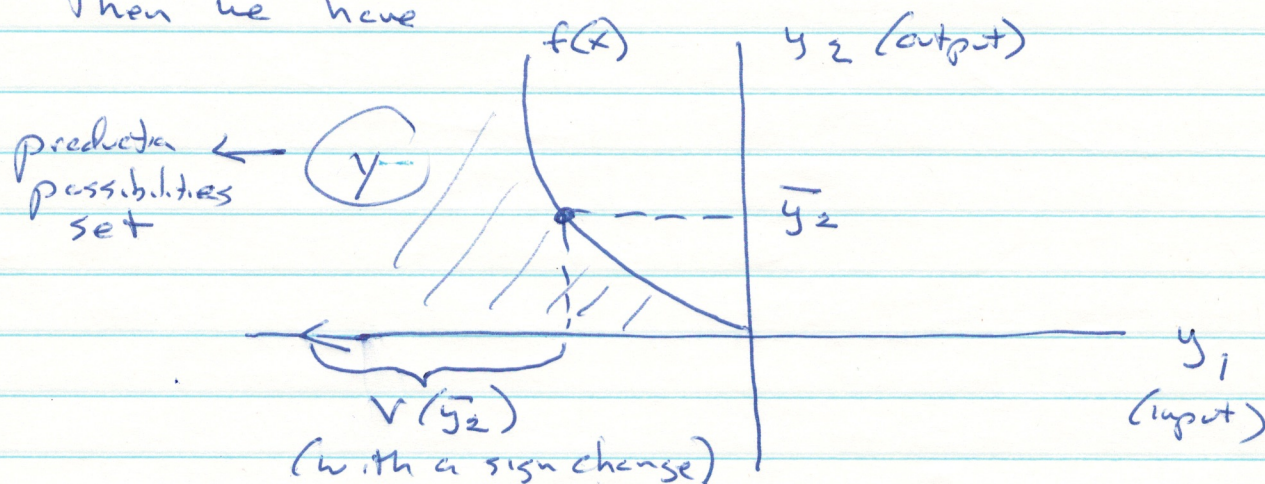
② However, even if all of the $V(y)$ sets are convex, this does not imply that Y is convex.

Here is a counter-example.

Suppose $(y_1, y_2) \in Y$ iff $0 \leq y_2 \leq y_1^2$

or $(y_1, y_2) \in Y$ iff $0 \leq y_1 \leq x^2$

Then we have



Clearly the set $V(\bar{y}_2)$ is convex; choose any two points along this part of the horizontal axis - then clearly the line segment between the points is in the set $V(\bar{y}_2)$.

But Y is not convex; we can choose two points along the boundary given by the quadratic production function $y_2 = y_1^2$ and the intermediate points will be outside the set Y .

③ $V(y)$ is convex iff $f(x)$ is a quasi-concave function.

By definition a function $f(x)$ is quasi-concave when $\{x : f(x) \geq y\}$ is a convex set for each y .

This is equivalent to saying that all of the input requirement sets are convex.

Why is quasi-concavity of the production function useful?

(a) we don't have to worry about second order conditions; if the first-order conditions for profit max hold, we automatically have a solution.

(b) If we have strict quasi-concavity (all of the input requirement sets are strictly convex), then any solution to a cost-minimization problem must be unique and the firm's conditional input demand functions will be nicely behaved (more on this later).

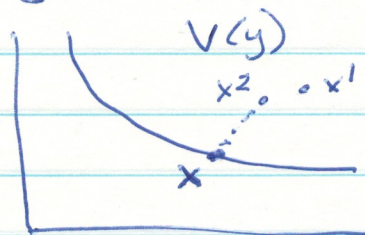
(3) Regularity (going back to my list of "reasonable" features of the production technology)
we say $V(y)$ is "regular" if it is a non-empty and closed set for all $y \geq 0$.

Non-empty just means that for any $y \geq 0$, there is at least one input vector x that will produce y .

Closed means that the set $V(y)$ includes its boundary. More technically, if you can find some sequence of input vectors $x^1, x^2, \dots, x^M, \dots$ that approach a vector x , and each member of the sequence is in $V(y)$, then $x \in V(y)$.

Graphically, we have something like this:

If $V(y)$ were not closed, the point x might not be in $V(y)$.



Relevance: This helps to ensure that cost-min problems such as $\min_{x \in V(y)} w \cdot x$

where $w = (w_1, \dots, w_n)$ is a price vector actually do have solutions; there will be some point in $V(y)$ at which cost is at least as low as it is at any other point (more on this later).

Now that we have discussed assumptions about input requirement sets, let's move on to a new topic.

Technical Rate of Substitution

This concept deals with the slope of the isoquant at a point. Varian explains this using two inputs. I will assume we have n inputs, just to be a bit more general:

Assume we have a production function $y = f(x)$.

The firm is currently producing $y^* = f(x^*)$.

We want to increase x_1 a bit, while decreasing x_n , to maintain the existing output y^* .

We will keep all other inputs fixed at current levels.

What is the rate at which x_n needs to be decreased as x_1 increases?

Note: we are moving along the isoquant; this is a question about its slope.

Define the function

$$x_n(x_1, \dots, x_{n-1}) \equiv \{x_n : f(x_1, \dots, x_{n-1}, x_n) = y^*\}$$

This function tells us how much x_n we need in order to produce y^* , given $x_1, \dots, x_{n-1}$ of the other inputs.

Assuming that the production function is differentiable and has positive first derivatives (all marginal products are positive) we can use the implicit function Theorem to show that $x_n(x_1, \dots, x_{n-1})$ exists and is differentiable in a neighborhood of x^* .

Now plug this function back into the production function to get

$$f(x_1, \dots, x_{n-1}, x_n(x_1, \dots, x_{n-1})) \equiv y^*.$$

This is an identity (it is true by definition). If you use the amount of x_n required to get y^* , then you will get y^* , regardless of the levels of $x_1, \dots, x_{n-1}$.

I will often use the notation " $\equiv$ " with three bars for an identity like this.

Now differentiate the identity on both sides with respect to x_1 and evaluate all derivatives at x^* . This gives

$$\frac{\partial f(x^*)}{\partial x_1} + \frac{\partial f(x^*)}{\partial x_n} \frac{\partial x_n(x_1^*, \dots, x_{n-1}^*)}{\partial x_1} = 0$$

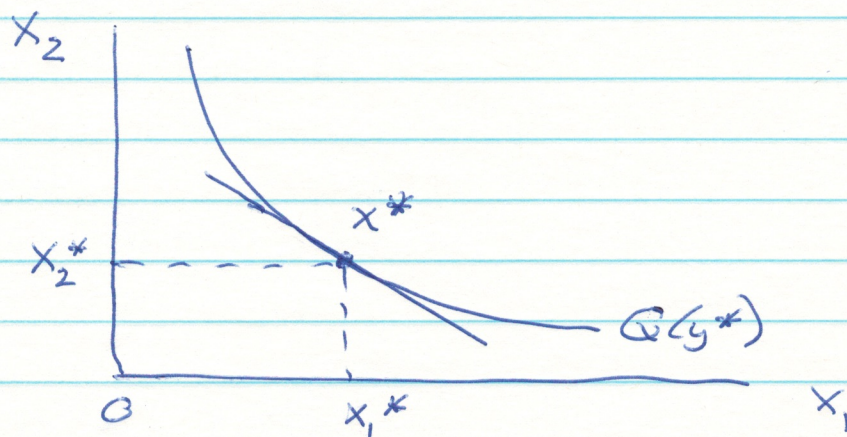
$$\text{or } \frac{\partial x_n(x_1^*, \dots, x_{n-1}^*)}{\partial x_1} = - \frac{\frac{\partial f(x^*)}{\partial x_1}}{\frac{\partial f(x^*)}{\partial x_n}} \equiv \text{TRS}$$

This is defined to be the technical rate of substitution between x_1 and x_n at x^* .

Notice that the TRS is the ratio of the marginal products with a minus sign in front. Because MP's are usually positive, in this definition TRS is negative.

Graphically:

At the point x^* on the isoquant $Q(y^*)$ TRS at x^* is equal to the slope of the tangent



line at x^* . This gives the rate at which x_1 and x_2 must be substituted in order to keep y^* constant.

Varian gives an algebraic example on p. 12 for a Cobb-Douglas production function.

The Elasticity of Substitution (an application of TRS)

The presentation of this concept by Varian on pp. 13-14 is hard to follow. I will try to explain it.

Definition:
$$\sigma = \frac{d(x_2/x_1)}{d|TRS|} \cdot \frac{|TRS|}{(x_2/x_1)} \quad \text{(assuming 2 inputs and 1 output)}$$

Comments: I added the absolute value sign for $|TRS|$ to convert it to a positive number. On p. 13, Varian doesn't use this but he slips into it on p. 14 when using logs, so let's just define σ that way to begin with.

What is the point of the concept?

When we get to cost minimization in later chapters, we will find that $|TRS| = \frac{\frac{\partial f(x_1^*, x_2^*)}{\partial x_1}}{\frac{\partial f(x_1^*, x_2^*)}{\partial x_2}} = \frac{w_1}{w_2}$

where $\frac{w_1}{w_2}$ is the input price ratio

and x^* is the cost-minimizing way to get some desired output y^* .

So then we will have $\sigma = \frac{d(x_2/x_1)}{d(w_1/w_2)} \cdot \frac{(w_1/w_2)}{(x_2/x_1)}$

which is something interesting: the elasticity of the input quantity ratio with respect to the input price ratio.

I will say later why it is useful to know about this. Unfortunately we are not ready yet to talk about cost minimization so Varian is treating σ as a purely technological concept, not a behavioral concept.

Why are we writing $\frac{x_2}{x_1}$ as a function of TRS?

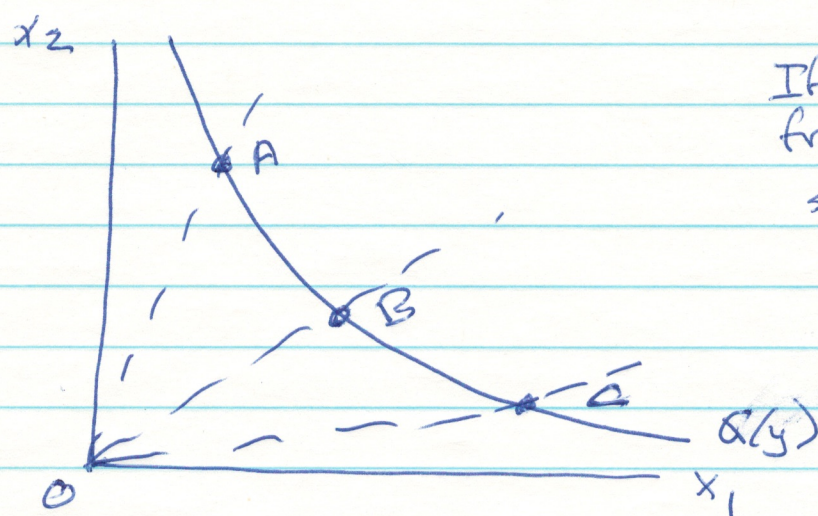
It does make sense in a cost-min context to think of $\frac{x_2}{x_1}$ as a function of $\frac{w_1}{w_2}$ through the firm's input demand functions. However, suppose we haven't heard of cost-min. How can we make sense of this?

It is true that $TRS(x_1, x_2) = - \frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}}$

so TRS is a function of (x_1, x_2) .

But it is not obvious from this that $\frac{x_2}{x_1}$ is a function of TRS.

However, think about it geometrically.

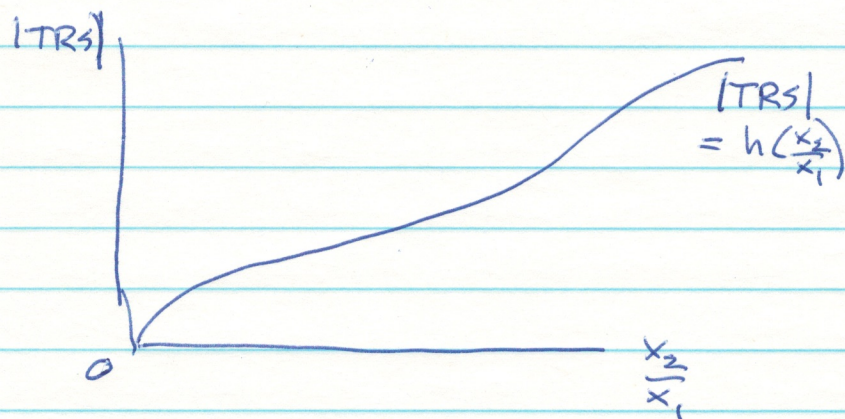


If the isoquant $Q(y)$ comes from a strictly convex $V(y)$ set, there will be a relationship between $\frac{x_2}{x_1}$ and TRS.

At point A, $\frac{x_2}{x_1}$ is large and TRS is also large (in absolute value)

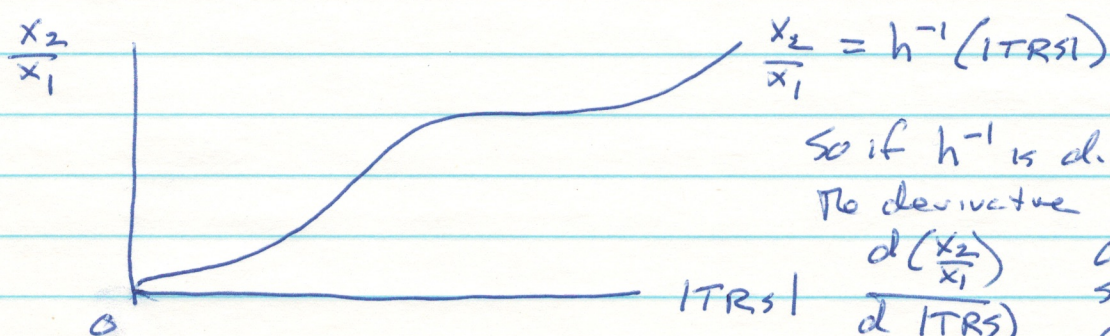
As we decrease $\frac{x_2}{x_1}$ by going to B and then C, the isoquant flattens out and $|TRS|$ drops.

~~Now invert~~ Call this function $|TRS| = h\left(\frac{x_2}{x_1}\right)$



We don't know much about the shape of this function but we know it is increasing.

Now invert this function and call the inverse h^{-1} .



So if h^{-1} is differentiable, the derivative $\frac{d\left(\frac{x_2}{x_1}\right)}{d|TRS|}$ does make sense and off we go.

That's all I want to say for now about the elasticity of substitution.

The last topic I want to discuss from Chapter 1 is returns to scale.

Returns to Scale

What happens to output when all inputs are changed in the same proportion?

This is a different question from convexity of $V(y)$ where we held output constant and discussed various input bundles that could produce the given output. It is also different from the TRS and elasticity of substitution, where we held output constant and moved along an isoquant.

There are three general situations that can arise

Constant Returns to Scale

We have CRS when $y \in Y$ implies $ty \in Y$ for any $t > 0$. This says you can take any feasible production plan y and scale it up or down in any way you want, as long as all inputs and outputs change by the same factor t . (So you could double everything, or multiply everything by $1/3$, etc.)

This not something we always want to assume but it is often convenient and at least somewhat reasonable.

The single-output versions are $x \in V(y) \Rightarrow tx \in V(ty)$, all $t > 0$
 $f(tx) = tf(x)$, all $t > 0$

CRS is related to the homogeneity of a function.
A function f is homogeneous of degree k if

$$f(tx) = t^k f(x) \text{ for all } t > 0$$

where t is a scalar.

This is not a necessary feature of production functions but if we happen to have a homogeneous production function with $k=1$ This implies CRS (often called linear homogeneity)

What is the rationale for CRS?

People sometimes argue that you can always replicate a feasible production plan by scaling it up in integer multiples (if you have one factory with a certain number of workers using some amount of raw materials to produce some amount of output, you should be able to create a second identical factory that does the same thing)

But there are some issues:

- ① you may not be able to scale things up in non-integer multiples, or subdivide them (indivisibilities)
- ② with more inputs, you may be able to do some things you couldn't do before, and get more than a proportional increase in output.
- ③ maybe the replication argument is irrelevant, because you can't get more of some inputs.

For these reasons, we want to consider other cases.

Increasing returns to scale: $f(tx) > tf(x)$ for all $t > 1$.

Decreasing returns to scale: $f(tx) < tf(x)$ for all $t > 1$.

Note that for both IRS and DRS, we are imposing the restriction $t > 1$, so we are only looking at what happens when we scale all inputs up simultaneously.

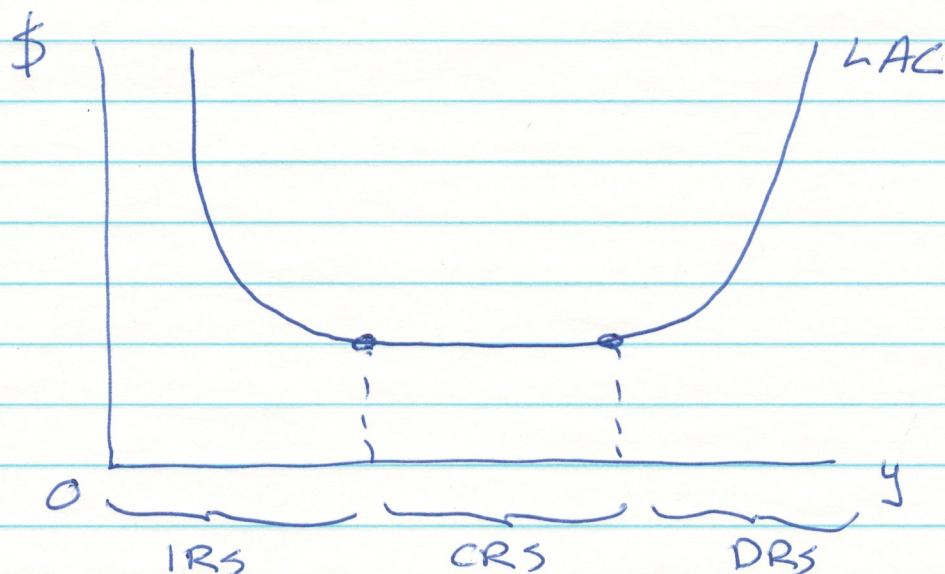
IRS $\Rightarrow$ get more than a proportional increase in output
 DRS $\Rightarrow$ get less " " " " " "

The idea behind IRS is that additional techniques may become feasible when you have more inputs. Also there may be engineering relationships that matter (for instance, the volume of a container grows more rapidly than its surface area).

The idea behind DRS is that you may be unable to increase certain inputs. This is typically true in short run situations where some inputs are fixed and you have diminishing returns when you increase the variable inputs. But it can also be true in the long run if there are inputs from nature that you don't control, like rainfall or sunshine.

Local returns to scale. CRS, IRS and DRS are all global concepts. But we can have production functions with CRS at some points, IRS at other points, and so on.

Suppose we have a U-shaped long run average cost curve, with a flat part at the bottom:



I will show later that we get decreasing LAC when the firm operates in a region of IRS, flat LAC when " " " " " of CRS and increasing LAC in a region of DRS.

Returns to scale at a point x are defined using the elasticity of output with respect to scale. At an input bundle x^0 , this is given by

$$\frac{\sum_{i=1}^n \frac{\partial f(x^0)}{\partial x_i} \cdot x_i^0}{f(x^0)}$$

or in vector notation $\frac{\frac{\partial f(x^0)}{\partial x} \cdot x^0}{f(x^0)}$
 (if $= 1$ then CRS;
 if > 1 , then IRS;
 if < 1 , then DRS)

I will derive this concept later and show how it can be used. That's all for Chapter 1.