DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
COLORADO STATE UNIVERSITY
MATH CAMP

NOTES
FALL, 2008

Lecture 1: Dr. C. Bond
Lecture 2: Dr. C. Goemans
Lecture 3: Dr. J. Bond
Lecture 4: Dr. D. Pendell
Lecture 5: Dr. C. Bond
Lecture 6: Dr. M. Costanigro

Lecture 1: Univariate Calculus and Economics
References: S&B Ch 1, Ch 3

INTRODUCTION

- In economics, we study the allocation of scarce resources, and this scarcity leads to tradeoffs in allocation decisions
- Mathematics is a *language* that we use to describe these tradeoffs
- In our most basic problems, the mathematics allows us to represent:
  - Consumer preferences and choice on the demand side (through utility functions)
  - Technology, costs, and profits on the production side (through production, cost, and profit functions)
  - Behavior (through maximization/minimization)
- Mathematics is extremely powerful in that it is:
  - Logical
  - Concise
  - Formal

  and can be used to tell economic stories about how economic agents are predicted behave under a set of assumptions
- The mathematical formalism in economics can be traced back to the 19th century "Marginal Revolution", with significant contributions by Leon Walras, Augustin Cournot, W. Stanley Jevons, and Alfred Marshall.
- All of the graphical analysis you have seen in your undergraduate economics education (demand, supply, indifference curves, marginal benefits/costs, OLS, etc…) is developed from formal mathematical analysis
- Just like the progression from Principles of Micro to Intermediate Micro (where you "got behind the scenes" of the main concepts like demand and supply), the next step in Graduate School is to progress into a deeper, more formal (read: mathematical) understanding of the relationships and decisions of economic agents and the assumptions behind how we model them.

CANONICAL FUNCTIONAL FORMS

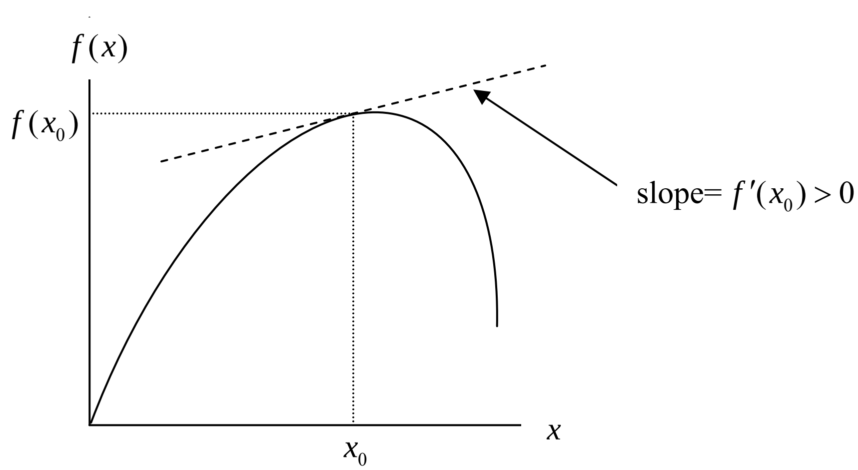
- A function is just like a "sausage machine"... you feed it one or more inputs (the arguments of the function, also called *exogenous* variables), and it returns one or more outputs (the value of the output once you feed in the inputs, or *endogenous* variables)
- A univariate function that maps from one (continuous) input into an output maps from the real line to the real line ($\Re^1 \to \Re^1$).
  - Canonical example: $y = f(x)$, where y is the dependent/endogenous variable and x is the independent/exogenous variable.
  - Specific example: $y = x^2$, where $f(x) = x^2$.
  - Economic examples:

- A utility function which maps the consumption of a good x to an ordinal measure of preferences: $v = u(x)$, where if $u(x_1) > u(x_2)$, then x_1 is preferred to x_2.
- A production function which maps the use of an input z into an output q: $q = f(z)$.

- Why use canonical forms?
    - Because they apply to a whole *class* of functions. In other words, we are representing *all* of the specific functions that admit the properties of the canonical function.

FIRST PARTIAL DERIVATIVES

- Recall that partial derivatives of a function approximate the change in the value of a function when the input changes by a very small amount.
- In other words, it is the *slope* of a function at the evaluation point, or the line just tangent to the function at a given evaluation point
- Formally, this derivative (if it exists) is defined by $\lim_{h_n \to 0} \frac{f(x_0 + h_n) - f(x_0)}{h_n}$.
- We want to know how to *interpret* a partial derivative and how to use them to *graph* economic relationships
- NOTATION: Let the partial derivative of a continuous, differentiable function $f(x)$ be denoted by as follows: $\frac{\partial f(x)}{\partial x} = f'(x) = f_x(x)$.
    - Specific example: $f(x) = x^2$, $f'(x) = 2x$.



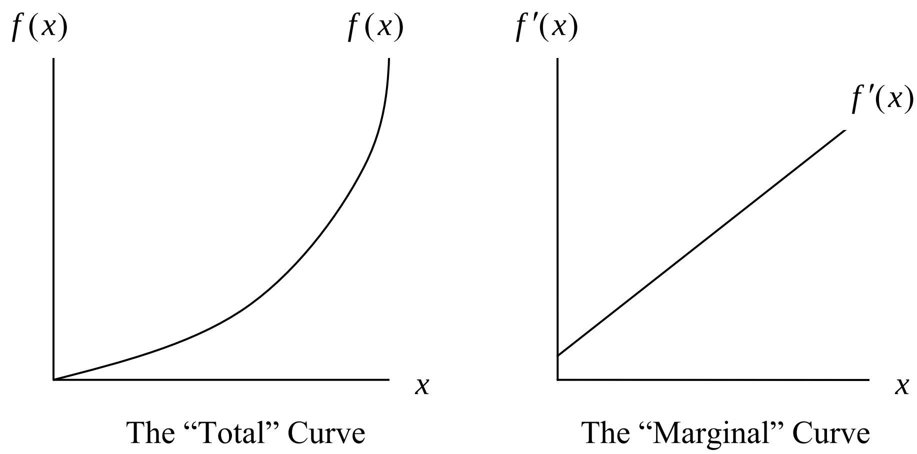
- In economics, the partial derivative plays a key role…it is the "*marginal*", or equivalently "*incremental*", effect of a change in the exogenous variable on the endogenous variable
  - Utility example: $u'(x)$ is how our utility changes with increases in the consumption of good x. This is called MARGINAL UTILITY. We would typically assume $u'(x) > 0$ for a "good", and $u'(x) < 0$ for a "bad", like pollution.
  - Production example: $f'(z)$ tells us how production changes with increases in input z. This is called MARGINAL PRODUCT. We would typically assume $f'(z) \geq 0$.
  - Note that in economic theory, it is often the case that a sign on a derivative is much more important than a magnitude

- PARTIAL DERIVATIVE = SLOPE OF A CURVE = "MARGINAL" IN ECONOMICS

SECOND PARTIAL DERIVATIVES AND GRAPHING

- In economics, we are usually concerned not only with first partial derivatives (see above), but second partial derivatives (the derivative of the first derivative).
- NOTATION: $\frac{\partial^2 f(x)}{\partial x^2} = f''(x) = f_{xx}(x)$.
- By natural extension, the second derivative is the "slope of the slope", or the slope of the marginal curve
- Economic examples:
  - $u''(x) < 0 \rightarrow$ decreasing marginal utility
  - $f''(z) < 0 \rightarrow$ decreasing marginal product
- As such, we can use the first and second derivatives of functions to provide a general graphical illustration of the "total" and "marginal" curves
  - The first partial derivative tells us the direction of the slope of the "total" curve and the level of the "marginal" curve
  - The second partial derivative tells us how the slope of the "total" curve changes and the slope of the "marginal" curve
- Example: Assume $f(x) \geq 0, f(0) = 0, f'(x) > 0, f''(x) > 0, \lim_{x \to \infty} f'(x) = \infty$.
  - The first assumption tells us that the value of the total curve $f(x)$ is always non-negative. The second assumption tells us that the value of $f(x)$ at $x = 0$ is zero.
  - The first partial derivative tells us that the total curve $f(x)$ is increasing in x and that the marginal curve $f'(x)$ is positive for all values of x.
  - The second partial derivative tells us that the slope of $f(x)$ is increasing with increases in x, and that the marginal curve $f'(x)$ has a positive slope.

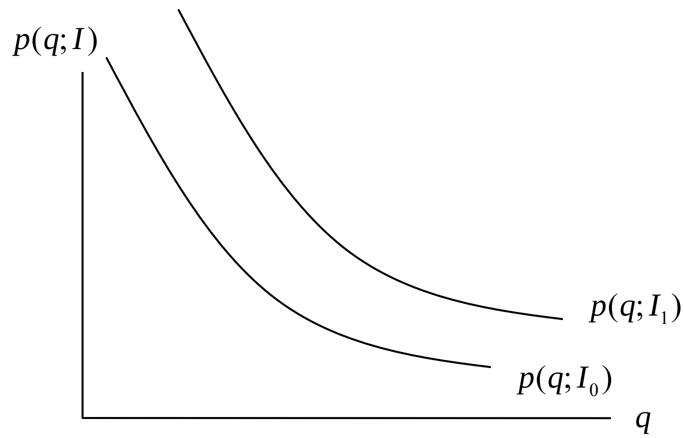
- The limit expression tells us that as x gets really large, the function $f(x)$ has an infinite slope, and the marginal curve $f'(x)$ has no limit.



- Note that without the third derivative, we cannot exactly represent how the slope of the marginal curve changes with changes in x.

A DERIVATIVE APPLICATION: ELASTICITIES

- Suppose we had a *demand function* $q(p;I)$, where q is quantity demanded of a certain good, p is the price, and I is a parameter representing income. Note that if we "inverted" this function to get $p(q;I)$, assuming we could, this is called the "*inverse demand function*".
  - Specific example: $q = \alpha + \beta p + \gamma I$.
- From a derivative standpoint, the law of demand suggests $q'(p;I) < 0$, or that as price increases, quantity demanded goes down. If this good was a normal good, this implies $\frac{\partial q(p;I)}{\partial I} > 0$, or demand increases as income increases. If it was an inferior good, then $\frac{\partial q(p;I)}{\partial I} < 0$.
  - For our specific example, this implies $\beta < 0$, $\gamma > 0$ for a normal good, and $\gamma < 0$ for an inferior good.



If $I_1 > I_0$, this is a normal good.

- Recall that economists like to summarize response to changes in a unit-free way…hence the idea of elasticities.
- Elasticities are defined by a formula such as: $\frac{\%\Delta Y}{\%\Delta X} = \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$.
- Note that if we make the ΔX very small, then $\frac{\Delta Y}{\Delta X} \approx \frac{\partial Y}{\partial X}$.
- As such, our own-price and income elasticities of demand become:

  - Own Price: $\frac{p \cdot q'(p;I)}{q(p;I)} < 0$,
  - Income: $\frac{I}{q(p;I)} \frac{\partial q(p;I)}{\partial I} \gtrless 0$.

DERIVATIVE RULES

Suppose that *k* is an arbitrary constant and that $f(x)$ and $g(x)$ are differentiable at the point $x = x_0$.

- If $z(x) = f(x) + g(x)$, then $z'(x_0) = f'(x_0) + g'(x_0)$
- If $z(x) = kf(x)$, then $z'(x_0) = kf'(x_0)$
- Product Rule: If $z(x) = f(x)g(x)$, then $z'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$
- Quotient Rule: If $z(x) = \frac{f(x)}{g(x)}$, then $z'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$
- Power Rule: If $z(x) = x^k$, then $z'(x_0) = kx_0^{k-1}$
- EXP: If $z(x) = e^x$, then $z'(x_0) = e^{x_0}$

- LN: If $z(x) = \ln(x)$, then $z'(x_0) = \dfrac{1}{x_0}$
- EXPONENTIALS: If $z(x) = b^x$, $b > 0$, then $z'(x_0) = \ln(x_0)b^{x_0}$

COMPOSITE FUNCITONS AND THE CHAIN RULE

- Composite functions just take a function of a function: $h(x) = g\left(f(x)\right)$.
    - Note that $h(x)$ is a function of x, and that $g(\cdot)$ has one argument. Our notation says: "take the output from the function $f(x)$ and use it as the argument, or input, into the function $g(\cdot)$.
    - Specific Example: $f(x) = x^{1/2}$ and $g(x) = x+1$, so $h(x) = g\left(f(x)\right) = x^{1/2} + 1$.
    - Economic example: The regression $y = \alpha + \beta\ln(x) + \varepsilon$, where $f(x) = \ln(x)$. This makes $g(z) = \alpha + \beta z + \varepsilon$.
- The chain rule for partial differentiation of a composite function is "the derivative of the outside function (evaluated at the inside function) times the derivative of the inside function.
    - Canonical: $h'(x) = g'\left(f(x)\right)\cdot f'(x)$.
    - Specific: $h'(x) = g'\left(f(x)\right) f'(x) = (1)\left(\dfrac{1}{2}x^{-1/2}\right) = \dfrac{1}{2}x^{-1/2}$.
    - Economic: $\dfrac{\partial y}{\partial x} = \beta / x$.

PRACTICE PROBLEMS

1. Assume a univariate utility function $u(x)$ that exhibits diminishing marginal utility in the good x.
    a. Graph the total and marginal utility functions on separate graphs.
    b. Confirm that the function $v(x) = x^{1/2}$ is a special case of $u(x)$.
    c. Our utility function is "homogeneous of degree one" if and only if $u(tx) = tu(x)$ for all *t*>0. Is $v(x) = x^{1/2}$ homogeneous of degree one? Explain.
    d. The economic concept of "strong monotonicity" in this context says that if $x_1 > x_2$, then x_1 is preferred to x_2. Does $u(x)$ represent preferences that are strongly monotone? Explain.
2. Consider a total cost function $c(x)$, which represents the total costs of producing output level x. Assume $c'(x) > 0$ and $c''(x) > 0$.
    a. Formally show that the marginal cost curve intersects the average cost curve $c(x)/x$ at the minimum of average cost.

b. Graph marginal costs and average costs on the same graph.
c. Assuming no fixed costs, the marginal cost curve is the supply curve for a competitive firm. Find an expression for the elasticity of supply for a firm with the cost function $c(x)$.

3. Consider the linear demand curve for $q(p\,;I)$ which follows the law of demand.
   a. Based on these assumptions alone, what mathematical properties can we ascribe to $q'(p\,;I)$? $q''(p\,;I)$? $\frac{\partial q(p\,;I)}{\partial I}$?
   b. Suppose that demand at price p_0 is elastic (i.e., the absolute value of the own-price elasticity of demand is greater than 1). Formally show that an increase in the price will decrease total expenditures.
   c. Suppose this good is normal and a luxury item. What does that tell us about $q'(p\,;I)$? $q''(p\,;I)$? $\frac{\partial q(p\,;I)}{\partial I}$?
4. Consider the function $y(x;\alpha) = \alpha + \beta \ln x$, where α and β are known, fixed constants.
   a. Graph this function if $\beta > 0$. Is this more likely to represent a demand or supply curve? Why?
   b. Graph this function if $\beta < 0$. Is this more likely to represent a demand or supply curve? Why?
   c. Give an expression for the elasticity of *y* with respect to *x*.
5. A logistic growth function that describes growth of a stock of a resource as a function of the stock is given by $G(x;r,k) = rx[1 - x/k]$, where r and k are positive parameters. Use the derivative properties of this function to graph the growth curve as a function of the stock, and the evolution of the stock over time starting at a relatively low initial level.
6. Formally show that the marginal revenue for a monopolist (who faces a downward sloping demand curve) is less than the price charged for any quantity greater than zero. Document any assumptions you make.
7. Show that $\frac{\partial \ln y}{\partial \ln x}$ gives the elasticity of *y* with respect to *x*.
8. Given $h(x) = g(f(x))$, sign $h'(x)$ and $h''(x)$ (if possible) under the following assumptions:
   a. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   b. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   c. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   d. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs

Lecture 2: Multivariate Calculus and Economics
References: S&B Ch 12,13, and 14; Chiang Ch 2 and 9

Lot of Ingredients Go into Sausage: Applying day 1 in a more complex world

- **Introduction:** In the "real world" decisions about production typically involve multiple inputs and/or outputs (e.g. it takes more than one good to produce something). Similarly, real world consumption decisions typically involve choosing between multiple goods. Extending our understanding of functions that "map" from more than one input to one (or more) outputs allows us to develop models that more accurately reflect the "real world".

  Examples of this include:
  - **Example 1:** Consumers getting utility from two goods x_1 and x_2
    - $u(x_1, x_2) = x_1 + x_2$

Reads: the total amount of utility/happiness derived from the consumption of goods x_1 and x_2 is equal to the total amount of x_1 I consume plus the total amount of x_2 I consume.

Questions we might want to ask:
1. How does this "function" relate to my actual preferences?
2. Do certain "types" of functions characterize certain "types" of preferences?

  - **Example 2:** Producers using multiple inputs to produce multiple outputs
    - $Y(L, K) = L^{\alpha} K^{\beta}$

Reads:

Questions we might ask:
1. How does this function relate to an actual production process?
2. Do certain types of function characterize certain types of production processes.

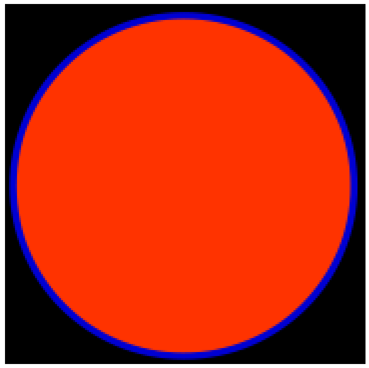
- **Some basics of Set Notation and Operation**
  - Economics is often about choosing between a "set" of goods, "inputs", "activities", etc. The term "set" simply refers to a collection of distinct objects
    - Example: The set of incoming graduate students
    - **Example:**

- Characterizing Sets
  - Enumeration
    - Example: $S_1 = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ or $S_2 = \{(1,3), (2,4), (8,5)\}$
    - **Example:**
  - Description
    - Example: $S_3 = \{(x, y) \mid x > 0, y > 0\}$
    - **Example:**
  - ***Question: What is the difference between ( ) and { }?***
- Set Membership
  - Example: $(-1, 20) \notin S_3$; $(1, 20) \in S_3$
  - **Example:**
- The complement of a set S in a universal set U is the set of all elements in U that are not in S and is denoted S^c
  - **Example:**
- Graphing Sets

- Denoting Relationships Between Sets
  - Set Equality: Set S is equal to set T if both sets contain identical elements
    - Example: $S_4 = \{1, 4, g, 9\}$ *and* $S_5 = \{4, g, 9, 1\}$
    - Example: The set of students taking this course "should" be equal to the set of new graduate students. Is this true?

  - Subsets: A set S is a subset of set T if every element of S is an element of T
    - Example: $S_6 \subset S_5$ where $S_6 = \{1, 4\}$ *and* $S_5 = \{4, g, 9, 1\}$
    - Example: The set of new graduate students is a subset of all graduate students at CSU.
    - Question: How many subsets can be formed from the set S_5
- Open, Closed and Compact Sets
  - Definition of an Open Set: $S \subset R^n$ is an open set if, for all $x \in S$, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset S$
    - Open ε-ball with center x^0 and radius $\varepsilon > 0$ is the subset of points in R^n:
      $B_\varepsilon(x^0) \equiv \{x \in R^n \mid \text{distance between } x^0 \text{ and } x \text{ is less than } \varepsilon\}$

- Definition of a Closed Set: S is a closed set if its complement, S^c, is an open set.
    - Example: The points (x,y) satisfying $x^2 + y^2 = r^2$ are colored blue. The points (x,y) satisfying $x^2 + y^2 < r^2$ are colored red. The red points form an open set. The union of the red and blue points is a closed set. Why?



- Definition of a Compact Set: A set **S** of real numbers is compact if and only if it is closed and bounded.

o "My care-factor towards sets is hovering within ε of zero." Why should I care? Weierstrass's Theorem :

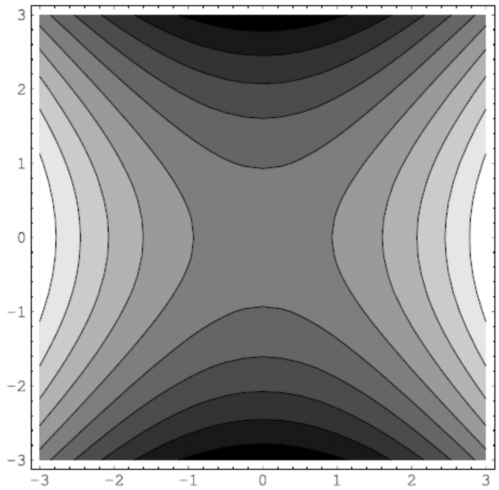
Let $S \subset R^n$ be compact, and let f: S→R be a continuous function on S. Then f attains a maximum and a minimum on S

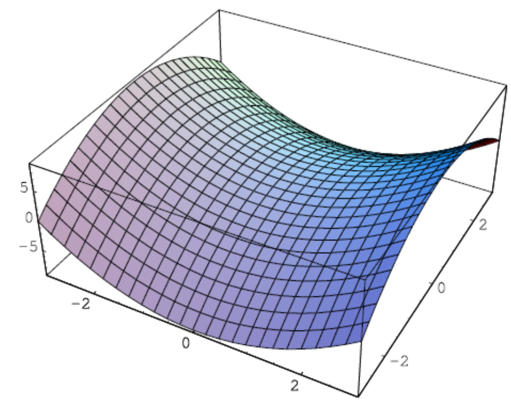
- Where does the maximum and minimum occur? Why do I care?

- Understanding Functions: Mapping from one set to another
  - Question: Doesn't S_1 look like a function?
    - A function is a set of ordered pairs with the property that any x value uniquely determines a y value.
  - $f : A \to B$; f is a function that takes elements from the set A and maps them into set B (I.e. sausage machine).
    - $f : R^1 \to R^1$ (from day 1)
      - Example: a production process that describes the use of one input to produce a single output
      - Example:
    - $f : R^n \to R^1$
      - Example: a production process that describes using n inputs to produce a single output
      - Example:
    - $f : R^n \to R^m$;
      - Example: a production process that describes producing m outputs from n inputs
      - Example:
  - Specific Examples:
    - Utility Chris derives from consuming different combinations of water and carrots
      - $u\left(carrots, water\right) = \alpha \ln\left(carrots + \phi\right) + \beta \ln\left(water\right)$
    - How many pies can I produce if I mix together x cups of flour, y cups of sugar, and z cups of butter?
      - $Y(flour, sugar, butter) = flour^{\alpha} suger^{\beta} butter^{\delta}$
- Graphing functions of two variables. Examples:
  - Level Curves: Let f be a function of two variables, and c a constant. The set of ordered pairs for which $f\left(x_1, x_2\right) = c$ is called the *level curve* for the value of c.
    - Example: Topographical maps?

Example

Graph $z = x^2 - y^2$.

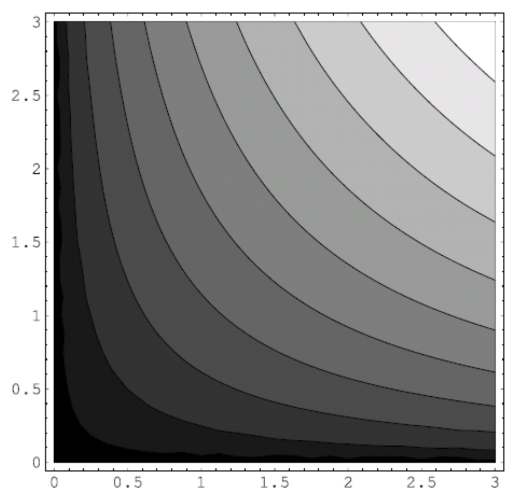


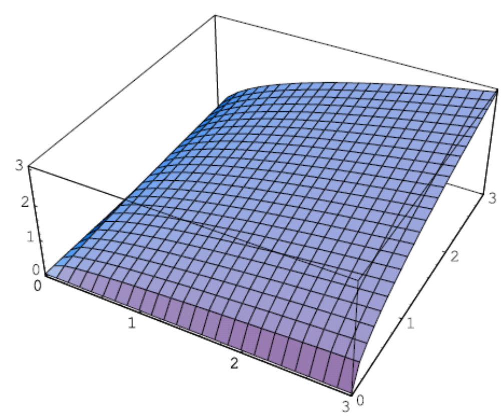


- Example: Indifference Curves

Example

Plot $z = x^{1/2}y^{1/2}$.

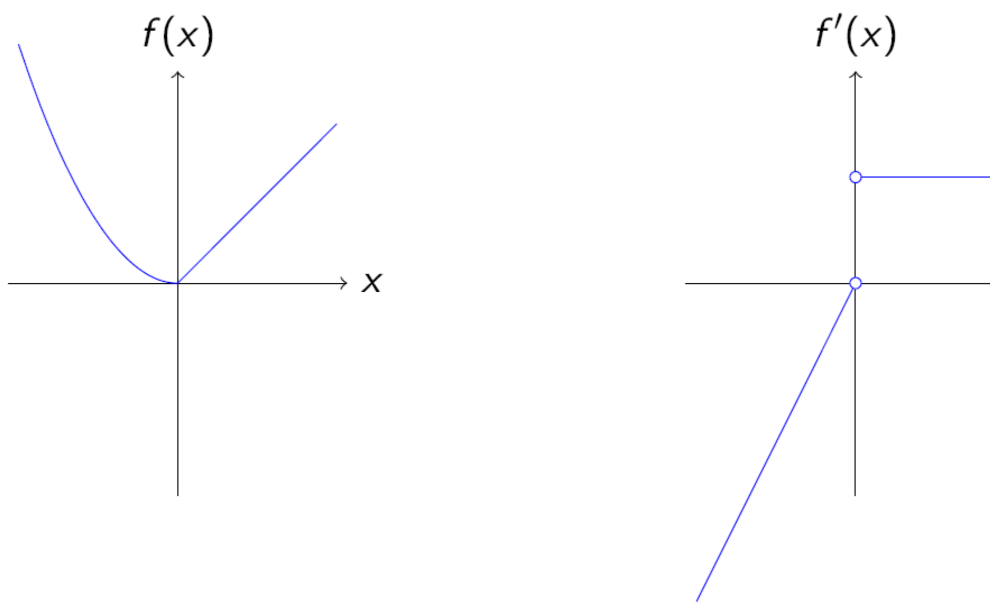




- Characterizing Functions (part 1)
  - Continuous Functions:
    - A function $f(x)$ is continuous at the point c (within the domain of f) if (1) $f(c)$ is defined and (2) $\lim_{x \to c} f(x) = f(c)$.

- A function $f(x_1,x_2)$ is continuous at the point $(\hat{x}_1,\hat{x}_2)$ (within the domain of *f*) if (1) $f(\hat{x}_1,\hat{x}_2)$ is defined and (2) $\lim_{(x_1,x_2)\to(\hat{x}_1,\hat{x}_2)} f(x_1,x_2) = f(\hat{x}_1,\hat{x}_2)$.
- A function $f(x_1,x_2)$ is continuous in x_1 at the point $(\hat{x}_1,\hat{x}_2)$ (within the domain of *f*) if (1) $f(\hat{x}_1,\hat{x}_2)$ is defined and (2) $\lim_{x_1\to\hat{x}_1} f(x_1,\hat{x}_2) = f(\hat{x}_1,\hat{x}_2)$.
- "Sharp points are OK, gaps are not."

- Continuity is an important concept because a function is differentiable only if it is continuous (i.e. continuity is a necessary condition).
  - Is it a sufficient condition? Why? Why not?



- Increasing Functions: Let $f: D \to R^n$, where D is a subset of R^n. Then *f* is increasing if $f(x^0) \geq f(x^1)$ whenever $x^0 \geq x^1$
  - Example: $f(x_1,x_2) = x_1 + x_2$
  - Example:
- Decreasing Functions: Let $f: D \to R^n$, where D is a subset of R^n. Then *f* is decreasing if $f(x^0) \leq f(x^1)$ whenever $x^0 \geq x^1$

- Example: $f\left(x_1, x_2\right) = -\left(x_1 + x_2\right)$
- Example:

- Partial and Total Derivatives
  - Partial Derivative : For $y = f(x_1, x_2)$ the partial derivative with respect to x_1 gives an approximation of the rate of change in y when there is a small change in x_1, holding constant x_2 (see S&B p. 300 for formal definition).
    - First Order Derivatives
      - $\dfrac{\partial f\left(x_1, x_2\right)}{\partial x_1} = f_{x_1}\left(x_1, x_2\right)$
      - Example: if $u\left(x_1, x_2\right) = x_1^{\alpha} x_2^{\beta}$ then $\dfrac{\partial u\left(x_1, x_2\right)}{\partial x_1} = \alpha x_1^{(\alpha-1)} x_2^{\beta}$
        - What does this tell us about the marginal utility of x_1? Does this make sense?
      - Example: Consider a production function $Q = Q(K, L)$
        - What is the derivative of this function with respect to K?
        - What does it mean if the derivative of this function with respect to K is positive? Negative?
      - Example:

    - Higher Order Derivatives
      - $\dfrac{\partial^2 f\left(x_1, x_2\right)}{\partial x_1 \partial x_1} = \dfrac{\partial^2 f\left(x_1, x_2\right)}{\partial x_1^{2}} = f_{x_1, x_1}\left(x_1, x_2\right)$
        - Example: if $u\left(x_1, x_2\right) = x_1^{\alpha} x_2^{\beta}$ then $\dfrac{\partial u\left(x_1, x_2\right)}{\partial x_1^{2}} = \alpha\left(\alpha - 1\right) x_1^{(\alpha-2)} x_2^{\beta}$
          - What does this tell us about the marginal utility of x_1? Does this make sense?
      - $\dfrac{\partial^2 f\left(x_1, x_2\right)}{\partial x_1 \partial x_2} = f_{x_1, x_2}\left(x_1, x_2\right)$
        - Example: if $u\left(x_1, x_2\right) = x_1^{\alpha} x_2^{\beta}$ then $\dfrac{\partial u\left(x_1, x_2\right)}{\partial x_1 \partial x_2} = \alpha\beta x_1^{(\alpha-1)} x_2^{(\beta-1)}$
          - What does this tell us about the marginal utility of x_1? Does this make sense?
    - Additional Examples:
      - Price Elasticity of Demand

            - Cross-Price Elasticity of Demand
- Characterizing Functions (part 2)
    - Continuously Differentiable Functions
    - (strictly) Concave and (strictly) Convex Functions
        - The function $y = f(x_1, x_2)$ is concave (convex) iff, for any pair of distinct points M and N on its graph- a surface- line segment MN lies either on or below (above) the surface. The function is strictly concave (strictly convex) iff line segment MN lies entirely below (above) the surface, except at M and N.

- Young's Theorem (Formal Def. S&B p. 330)
    - $\frac{\partial f(x,y)}{\partial x \partial y} = \frac{\partial f(x,y)}{\partial y \partial x}$
    - How is this relevant to the examples presented earlier?
- Total Derivative
    - Def. S&B p. 308)
    - Difference between partial and total derivative

Lecture 3: Unconstrained Optimization
References: S&B Ch 17, 22

Introduction:

- Economic theory is largely based on optimization theory. (An optimization problem is a mathematical problem in which one is either maximizing or minimizing the value of some objective function.)
- The two principle agents in micro theory are consumers and firms. At the most basic level, we assume that consumers maximize their utility while firms are assumed to maximize profit.
- Another key economic concept that evokes optimization is that of economic efficiency (e.g. a given objective can be achieved at a minimum cost).
- The two basic types of economic optimization problems are 1) constrained and 2) unconstrained. In unconstrained optimization problems one is trying to find the values of the choice variables that max or min some objective function, $f(x_1, x_2, ...x_n)$. In the next lecture you will learn about constrained optimization in which one is still attempting to max or min the objective function $f(x_1, x_2, ...x_n)$ but subject to one or more constraints on the choice variables (equalities, inequalities).

Generalized Univariate Example (Framework for Profit Max)

- Firm chooses level of output y to max profits.
- $\pi(y) = r(y) - c(y)$ where $\pi(y) =$ profit, $r(y) =$ total revenue, $c(y) =$ total costs
- One specific case is $\pi(y) = p * y - c(y)$
- To solve, find the first order necessary conditions (*FONC*). The FONC is found by taking a derivative of the profit function with respect to the choice variable – in doing so we are performing *marginal analysis*.
- Why necessary? At optimum, we want *FONC* to equal 0, (e.g., Marginal Revenue-Marginal Cost=0), at this point there is no net benefit from producing an additional unit of output. (Figure a)
- *FONC* $\frac{\partial \pi}{\partial y} = \pi'(y) = p - c'(y) = MR - MC$
- MR=MC means that the extra revenue gained from selling one additional unit of output just equals the extra cost of producing one more unit. An equivalent statement is that marginal profits are zero.
- The *FONC* implicitly define the optimal level of output (*y)* as a function of price. This *y** can be substituted into the profit function to get max profits for a given level of p.

- Next we check the second order sufficient conditions *SONC*, the *SONC* < 0
- Why sufficient? Because the slope of the marginal cost curve must be upward sloping (more generally, marginal profit must be decreasing)

- $SONC \frac{\partial^2 \pi}{\partial y^2} = \pi''(y) = -c''(y) < 0$, so the slope of the marginal cost curve, $c''(y) > 0$. See Figure a.

Univariate Numeric Example:

- Consider the case of a competitive firm which sells its output for a constant price *p* and whose optimization problem can be described as below. Using the information below, we will find the firm's supply curve and determine how quantity supplied changes with price and how supply changes with changes in marginal cost.
- $\pi(y) = p * y - \frac{1}{2}ky^2$ where $\pi(y) =$ profit, and total costs, $c(y) = \frac{1}{2}ky^2$
- $FONC \frac{\partial \pi}{\partial y} = \pi'(y) = p - ky = 0$ and $\therefore \; p = ky$
- Solve for optimal level of output, $y^* = \frac{p}{k}$

Related Comparative Statics

- We calculate comparative statistics to examine changes in equilibrium response to a change in underlying economic parameters.
- Comparative: because we are comparing one statistic in equilibrium to the new one that would occur if the parameters were to change.
- Static: because we do not describe the dynamic path of how the equilibrium actually moves from one position to another.
- Mathematically, we are taking the derivatives of our solution function wrt the parameters.
- Recall the previous example where $y^* = \frac{p}{k}$, now see how output would change with an change in price: $\frac{\partial y^*}{\partial p} = \frac{1}{k} > 0$, in this case an increase in price leads to an increase in quantity supplied. See Figure b.
- Now consider and increase in input costs: $\frac{\partial y^*}{\partial k} = -\frac{p}{k^2} < 0$, as costs for the supplier increase, amount supplied goes down. See Figure c.

Multivariate Numeric Example:

- Consider a simple monopoly price discrimination problem where the monopolist sells its product in two separate markets with demand curves given by: $q_1 = 100 - p_1$ and $q_2 = 60 - p_2$. The monopolist will generally want to charge different prices to the two groups and assume that re-selling of the good by consumers is not possible. Let the monopolist's total cost function be defined as: $C = 1000 + 20(q_1 + q_2)$.

- The monopolist's goal is to max profits (π) and the objective function is: $\pi = p_1q_1 + p_2q_2 - 1000 - 20(q_1 + q_2)$. Using the demand functions to solve for q_1 and q_2 in terms of p_1 and p_2. Profit expressed as a function of p_1 and p_2 is: $\pi(p_1, p_2) = 120p_1 + 80p_2 - p_1^2 - p_2^2 - 4200$.
- The first order conditions for maximizing profit are: $\frac{\partial \pi}{\partial p_1}\pi(p_1, p_2) = 120 - 2p_1 = 0$ and $\frac{\partial \pi}{\partial p_2}\pi(p_1, p_2) = 80 - 2p_2 = 0$. Solving for the critical values of p_1 and p_2 results in $p_1^* = 60$ and $p_2^* = 40$.

Econometrics Application: Minimizing the Sum of Squared Errors (MSSE)

- Similar to our unconstrained profit max problems; in the following case we are using a tool (regression analysis) to examine the relationship of a dependent variable to independent or explanatory variables. In the case of a profit maximizing firm we are trying to find output as a function of price.
- In a regression we are trying to determine what is a "good fit" (e.g. which estimated parameters best approximate the true relationship of the variables). To make this determination, we minimize the sum of squared errors. Doing so has several advantages over competing methods 1) avoids problems of sign 2) regression line goes through the middle point 3) squaring emphasizes large errors 4) method is easily manageable 5) has a unique minimum and 5) has a unique, and best, solution.
- Mathematically, MSSE is represented by: $\min \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n}\varepsilon^2$
- A simple regression model: $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$
- Where $\hat{\alpha}$ and $\hat{\beta}$ are estimates of the true, but unknown parameters α and β
- To minimize the sum of squared errors, $\min \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n}(y_i - \hat{\alpha} + \hat{\beta}x_i)^2$
- At optimum, $\frac{\partial}{\partial \hat{\alpha}}\sum_{i=1}^{n}\varepsilon_i^2 = 0$
- And, $\frac{\partial}{\partial \hat{\beta}}\sum_{i=1}^{n}\varepsilon_i^2 = 0$
- Where, $\hat{\beta} = \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$ and $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$

Practice Problems:

1. A simple monopolist faces a linear inverse demand curve given by $p(y) = a - by$ where $a > 0$ and $b > 0$, and where y is the output of the monopolist. In addition, the monopolist faces the total cost function given by $c(y) = ky$, where $k > 0$. Furthermore, assume that $a > k$.
   a. Set up the profit maximization problem facing the monopolist

b. Provide both graphical and economic interpretations of the assumption $a > k$. It may help to provide an economic interpretation of parameters a and k individually first.
c. What does the assumption $a > k$ imply for the profit maxing output level of the monopolist? Use the graph in part b to answer this question.
d. Find the monopolist's profit maximizing level of output, say $y = y^m(a,b,k)$, as well as its profit maximizing price, say $p = p^m(a,b,k)$. How do you know that you have found the profit maximizing solution?
e. Derive the comparative statistics, $\frac{\partial y^m}{\partial k}$ and $\frac{\partial p^m}{\partial k}$
f. Provide an economic interpretation of the above comparative statistics.

2. In this question, we will examine the effect of a *sales tax* on a monopolist's optimal choice of price and output. Assume that the monopolist faces a linear inverse demand curve given by $p(y) = a - by$ where $a > 0$ and $b > 0$, and a quadratic cost function given by $c(y) = \frac{1}{2}ky^2$, where $k > 0$. The monopolist is also assumed to pay a fraction s, $0 < s < 1$, of its total sales revenue in taxes to the government.
   a. Set up the monopolist's profit maximization problem in the presence of the sales tax, and find the profit max level of output, say $y = y^s(a,b,k,s)$.
   b. Find the monopolist's optimal price, $p = p^s(a,b,k,s)$.
   c. Derive the comparative statistics for an increase in the sales tax, namely $\frac{\partial y^s}{\partial s}$ and $\frac{\partial p^s}{\partial s}$, and provide an economic interpretation of these statistics
   d. What is the economic interpretation of an increase in a? You may want to plot a demand function to help with your interpretation.
   e. Derive the comparative statics for an increase in the parameter a, namely $\frac{\partial y^s}{\partial a}$ and $\frac{\partial p^s}{\partial a}$, and provide an economic interpretation of these statistics.
   f. What is the economic interpretation of an increase in the parameter k?
   g. Derive the comparative statics for an increase in the parameter a, namely $\frac{\partial y^s}{\partial k}$ and $\frac{\partial p^s}{\partial k}$, and provide an economic interpretation of these statistics.

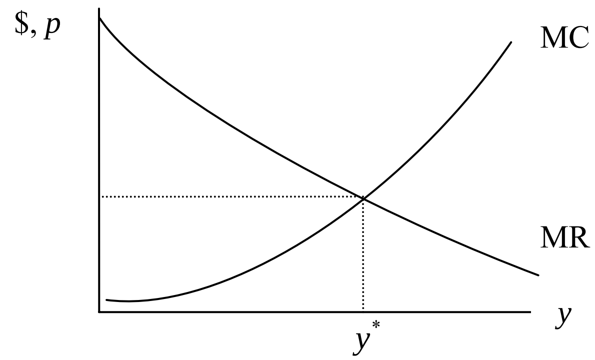


Figure a: Marginal Cost and Marginal Revenue Curves

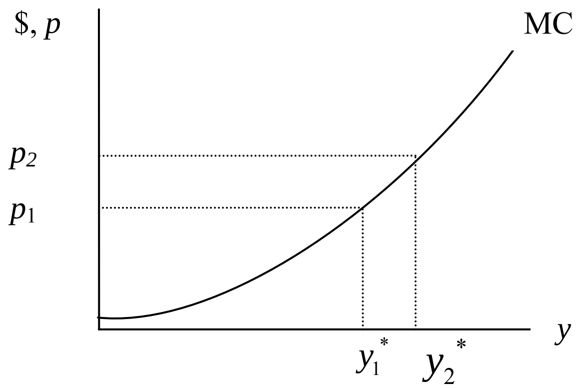


Figure b: Marginal Cost Curve with output price increase

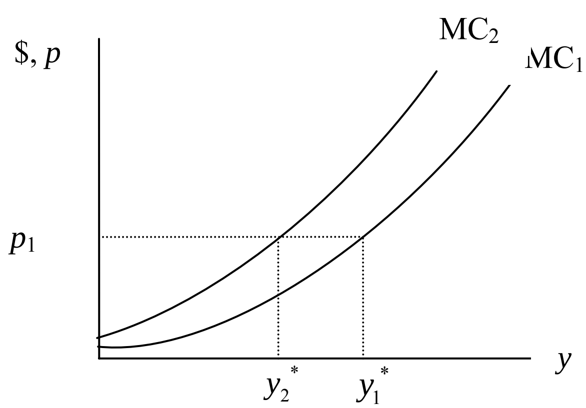


Figure c: Marginal Cost curve with input cost increase

Lecture 4: Constrained Optimization
References: S&B Ch 18, 19, 22

Economics is by large a science of choice. When an economic project is to be carried out, there are normally several alternative ways to accomplishing it. The most common criterion of choice among alternatives in economics is the goal of *maximizing* something (such as maximizing a firm's profit or a consumer's utility) or *minimizing* something (such as minimizing total costs). As discussed earlier, we may categorize such maximization and minimization problems under the general heading of *optimization*.

In formulating an optimization problem, we first need to define an *objective function* in which the dependent variable represent the object of maximization or minimization and in which the set of independent variables (or choice variables) indicates the objects which magnitudes the economic unit in question can pick and choose, with a view to optimizing. The essence of the optimization process is simply to find the set of values of the choice variables that will yield the desired extreme of the objective function.

Economists define models that represent the choices made by agents. These models are mathematical representations of the trade-offs inherent in decision process which provide tractable methods to solve for optimal behavior. These models enable derivation of testable hypotheses that should hold if model accurately reflects behavior. There are two types of optimization: (1) set of choices may be unlimited – unconstrained optimization (as discussed by Jennifer Bond) & (2) set of choices may be limited by budget constraints, resource scarcity, etc. – constrained optimization.

Consider an economic agent that receives benefit which can be represented by a function *f*(x). The function *f* is called the objective function. *f*(x) could represent profits or utility and *x* is the quantity produced or good consumed. The objective of such an agent is to select *x* so as to maximize benefits (or in the case of cost, expenditure, or risk, the agent would select *x* for minimization).

In many real world applications, the set of feasible choices is constrained (individuals have a finite budget set to spend on purchases; factors of production are finite and scarce). Let's look at an example.

$$\max_{x} \; f(x)$$

subject to: $g(x) \le b$ and $h(x) = c$

Where *f* is the objective function, *x* is an input or good, *g* and *h* are the constraint functions. Specifically, *g* is an *inequality* constraint and *h* is an *equality* constraint. In many applications, the most common inequality constraints are nonnegativity constraints: $x \ge 0$ while equality constraints often are definitions of one variable in terms of another (e.g., budget constraint).

Examples

[Point out objective functions, choice variables, inequality and equality constraints.]

Utility Maximization Problem – Example 1

In the most basic problem x_i represents the amount of good *i* and $f(x_1, x_2,\ldots, x_n)$, usually written at $U(x_1, x_2,\ldots, x_n)$, measures the individuals' level of utility or satisfaction with consuming x_1 units of good 1, x_2 units of good 2, and so on. Let $p_1, \ldots, p_n$ denote the price of the goods and let *M* denote the individuals income. The consumer wants to

Maximize $U(x_1, x_2,\ldots, x_n)$

Subject to $p_1x_1 + p_2x_2 +\ldots+ p_nx_n \leq M,$

$x_1 \geq 0, x_2 \geq 0,\ldots x_n \geq 0.$

Utility Maximization Problem with Labor/Leisure Choice – Example 2
Let *U, M,* $x_1, x_2,\ldots, x_n$, and $p_1, \ldots, p_n$ be as in the preceding example. In addition, Let *w* denote the wage rate, *M'* denote the consumer's nonwage rate income, l_0 as hours of labor, and l_1 as the hours of leisure. The consumer has $M' + w\, l_0$ dollars to spend and wants to

Maximize $U(x_1, x_2,\ldots, x_n, l_1)$

Subject to $p_1x_1 + p_2x_2 +\ldots+ p_nx_n \leq M' + w\, l_0,$

$l_0 + l_1 = 24,$

$x_1 \geq 0, x_2 \geq 0,\ldots x_n \geq 0, l_0 \geq 0, l_0 \geq 0.$

Profit Maximization of a Competitive Firm – Example 3

Suppose that a firm in a competitive industry uses *n* inputs to manufacture its products. Let *y* denote the amount of its output, and let $x_1, x_2,\ldots, x_n$ denote the amounts of its inputs. Let $y = f(x_1, x_2,\ldots, x_n)$ denote a firm's production function, describing the maximal amount of output that can be produced from the inputs bundle $(x_1, x_2,\ldots, x_n)$. Let *p* be the units price of the output and let w_i denote the cost of input *i*. The firm's goal is to choose $(x_1, x_2,\ldots, x_n)$ to maximize its profit:

$\prod (x_1, x_2,\ldots, x_n) = p^* f(x_1, x_2,\ldots, x_n) - \sum w_ix_i$

Under the constraints

$p^* f(x_1, x_2, \ldots, x_n) - \sum w_i x_i \geq 0$ (reflects the requirement that the firm make a non-negative profit),

$g(x) \leq b_1, \ldots, g(x) \leq b_k$ (availability of the inputs),

$x_1 \geq 0, \ldots, x_n \geq 0$ (non-negative inputs).

Let's revisit the basic consumer problem and principle/intermediate microeconomic class.

We have two goods, x_1 and x_2. Recall from you microeconomics class, we can graph the indifference curves (see Figure 1).[1] What do we know given Figure 1?

Point A (or bundle A) has more x_2 than point B, but B has more X_1 than point A. Would the individual prefer point A or point B? Why? We know the individual equally likes B and C. Also, we know that the individual prefers C to A. We can infer that she likes B more than A.

At what point will the consumer maximize their utility? Why? At point C the indifference curve is tangent to the budget constraint (in other words the slopes are equal).

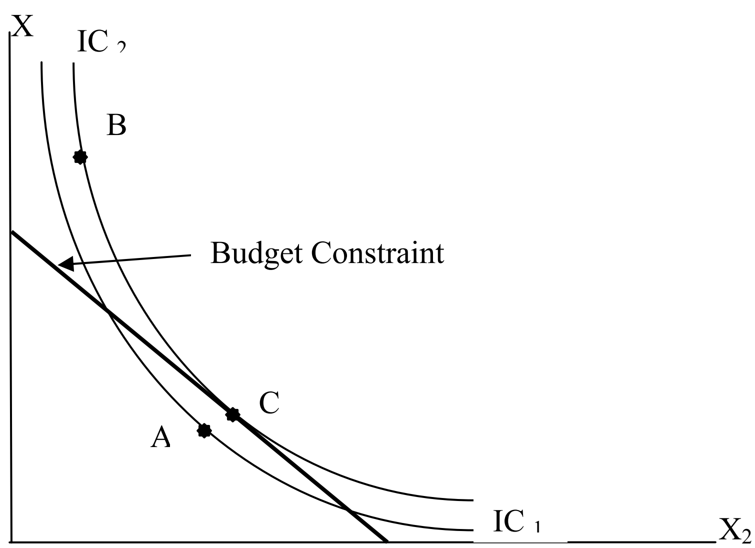


Figure 1.

[1] An indifference curve is different bundles or baskets of goods which a consumer considers equally desirable.

Let's relate what you have discussed in previous lectures (by Craig, Chris, and Jennifer) to the figure.

Figure 1 highlights that the shape of the indifference curve is not a straight line. It is conventional to draw the curve as bowed. Why? Because of the concept of diminishing *marginal rate of substitution* (MRS) between the two goods. The MRS between two goods is the slope of the indifference curve for the two goods. In other words, the rate at which a customer is ready to give up one good in exchange for another good while maintaining the same level of satisfaction.

How do we calculate MRS? Recall, the MRS = change in good X_2 / change in good X_1. Can we use partial differentials to calculate this? Why not? The partial derivative of a function is the derivative w.r.t. one of the variables while the others are held constant. Instead, we need to think about this using the total derivative. The total derivative of a function, *f*, of several variables, e.g., *x,y,z* etc., with respect to one of its input variables, e.g., *x*, is different from the partial derivative. Calculation of the total derivative of *f* with respect to *x* does not assume that the other arguments are constant while *x* varies; instead, it allows the other arguments to depend on *x*. For example, the total derivative of *f(x,y,z)* w.r.t. to *x* is:

$$\frac{df}{dx} = \frac{df}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} + \frac{\partial f}{\partial z}\frac{dz}{dx}$$

Recall from Lecture 1, the concept of marginal utility. What is marginal utility?

Marginal Utility (MU) is the amount of additional utility an individual acquires from an additional unit of a good when the amount of another good is held constant. The marginal utility of x_1 (MU_{x1}) is amount of additional utility an individual acquires from an additional unit of x_1 when the amount of x_2 is held constant. MU_{x2} is similarly defined.

How do we calculate Marginal Utility?

As discussed by Craig (in Lecture 1), the first partial derivative of the consumer's utility function.

We know from intermediate micro class that the ratio of the marginal utilities are equal to the MRS. In other words, $MRS_{x1x2} = MU_{x1} / MU_{x2}$. How do we know that?

Assume the following consumer utility function: $U(x_1, x_2)$.

Prove the marginal rate of substitution of x_1 for x_2 (i.e., MRS_{x1x2}) is equal to the ratio of the marginal utilities (i.e., MU_{x1} / MU_{x2}).

Marginal Utility:

$$MU_{x1} = \frac{\partial U(\cdot)}{\partial x_1}$$

$$MU_{x2} = \frac{\partial U(\cdot)}{\partial x_2}$$

Marginal Rate of Substitution:
Take the total differential of the utility function equation, we obtain the following results:

$$dU(\cdot) = \left(\frac{\partial U(\cdot)}{\partial x_1}\right) dx_1 + \left(\frac{\partial U(\cdot)}{\partial x_2}\right) dx_2$$

Substituting MU_{x1} and MU_{x2}:

$$dU(\cdot) = \left(\frac{\partial U(\cdot)}{\partial x_1}\right) dx_1 + \left(\frac{\partial U(\cdot)}{\partial x_2}\right) dx_2$$
$$= MU_{x_1} dx_1 + MU_{x_2} dx_2$$

Divide both sides by dx_1:

$$\frac{dU(\cdot)}{dx_1} = \frac{MU_{x_1} dx_1}{dx_1} + \frac{MU_{x_2} dx_2}{dx_1}$$
$$= MU_{x_1} + MU_{x_2} \frac{dx_2}{dx_1}$$

Through any point on the indifference curve, $\frac{dU(\cdot)}{dx_1} = 0$, because $U = c$ (where c is a constant). Thus,

$$0 = MU_{x_1} + MU_{x_2} \frac{dx_2}{dx_1} \Rightarrow$$

$$\frac{MU_{x_1}}{MU_{x_2}} = -\frac{dx_2}{dx_1}$$

We have proven the ratio of the marginal utilities equals the slope of the indifference curve.

Now let's incorporate the constraint in the problem. Recall, we are trying to maximize the utility given a budget constraint ($p_1x_1 + p_2x_2 \leq M$). The budget line (or budget constraint – Figure 1) illustrates all the possible combinations of two goods that can be purchased at given prices and for a given consumer budget. How do we calculate the slope of the budget line? The slope of the budget line is the ratio of the price of X_1 to the price of X_2. A rational, maximizing consumer would prefer to be on the highest possible indifference curve given their budget constraint. This point occurs where the indifference curve touches (is tangential to) the budget line. In the case of Figure 1, the optimum consumption point occurs at point C on indifference curve IC_2 ($MRS_{x1x2} = MUx_1 / MUx_2 = P_{x1} / P_{x2}$).

Revisiting the consumer problem.

Maximize $U(x_1, x_2)$

Subject to $p_1x_1 + p_2x_2 = M$

Marginal Utility =

$$MU_{x1} = \frac{\partial U(\cdot)}{\partial x_1}$$

$$MU_{x2} = \frac{\partial U(\cdot)}{\partial x_2}$$

What is the slope of the indifference curve? $\dfrac{\frac{\partial U(\cdot)}{\partial x_1}}{\frac{\partial U(\cdot)}{\partial x_2}} = \dfrac{MU_{x1}}{MU_{x2}}$

What is the slope of budget constraint? – p_1 / p_2

If, $\dfrac{MU_{x1}}{MU_{x2}} = \dfrac{p_1}{p_2}$, then we are at point C in the graph.

Lagrangian Method

How do economists model choices when some set of actions are constrained? We employ tools of constrained optimization – Lagrangian which incorporate bounds on set of feasible actions.

Maximize $U(x_1, x_2)$

Subject to $p_1x_1 + p_2x_2 = M$

One method is to solve the budget constraint for X_1 and then substitute it in the objective function. Solving by substitution is easy when you have linear constraints and a small number of variables and constraints. Another method commonly employed in economics is the Lagrange method.

Formally, we introduce the income constraint into the problem as follows:

$$L\ (x_1, x_2, \lambda) = U(x_1, x_2) + \lambda\left(M - p_1x_1 - p_2x_2\right)$$

where, λ is the Lagrangian multiplier.

You will notice we've incorporated the constraint and created an unconstrained maximization problem, with three variables: x_1, x_2, and λ (lambda), which is called the Lagrange Multiplier.

Let's find the first order conditions (FOCs). How do we do that? We will partially differentiate the Lagrangian w.r.t. the three variables.

$$1)\ \frac{\partial L(\cdot)}{\partial x_1} = \frac{\partial U\ (\cdot)}{\partial x_1} - \lambda\ p_1 = 0$$

$$2)\ \frac{\partial L(\cdot)}{\partial x_2} = \frac{\partial U\ (\cdot)}{\partial x_2} - \lambda\ p_2 = 0$$

$$3)\ \frac{\partial L(\cdot)}{\partial \lambda} = M - p_1x_1 - p_2x_2 = 0$$

Note that partially differentiating the Lagrangian with respect to λ will produce the original constraint. [What if you have more than one constraint?]

Using the first two FOCs, we can obtain:

$$1)\ \frac{\partial U\ (\cdot)}{\partial x_1} = \lambda\ p_1 \Rightarrow \frac{\frac{\partial U\ (\cdot)}{\partial x_1}}{p_1} = \lambda$$

$$2)\ \frac{\partial U\ (\cdot)}{\partial x_2} = \lambda\ p_2 \Rightarrow \frac{\frac{\partial U\ (\cdot)}{\partial x_2}}{p_2} = \lambda$$

Rearrange,

$$\frac{\frac{\partial U\ (\cdot)}{\partial x_1}}{p_1} = \lambda = \frac{\frac{\partial U\ (\cdot)}{\partial x_2}}{p_2} \Rightarrow \frac{\frac{\partial U\ (\cdot)}{\partial x_1}}{p_1} = \frac{\frac{\partial U\ (\cdot)}{\partial x_2}}{p_2} \Rightarrow \frac{\frac{\partial U\ (\cdot)}{\partial x_1}}{\frac{\partial U\ (\cdot)}{\partial x_2}} = \frac{p_1}{p_2}$$

That is, the individual selects the point on the budget constraint at which the ratio of the marginal utilities is equal to the price ratio.

How do we verify we have found the maximum? We need to verify the second-order condition.

Why bother with the Lagrangian method? First, as was already mentioned, in many problems the Lagrange multiplier has nice interpretations that allow for additional intuition. Second, the process of solving an optimization problem using the Lagrange method provides additional insights that cannot be recovered from the substitution method. Third, solving more complex optimization problems is facilitated by using this method. Finally, it is so widely used within economics.

Lambda (or the Lagrangian Multiplier). The Lagrange multiplier, λ, represents the effect of a small change in the constraint on the optimal value of the objective function. What does lambda imply in the consumer maximization problem? Suppose we are given one additional unit of income (M) (i.e., we marginally relax the constraint). What happens to the value of the objective function? In other contexts, the Lagrange multiplier may be interpreted differently. For example, if the objective function represents the profit function from undertaking an activity and the constraint reflects a limit on using an input, the Lagrange multiplier reflects the marginal benefit from having additional input. In this case the Lagrange multiplier represents the price a firm would be willing to pay per unit of additional input, which is known as the **shadow price** of the input.

The envelope theorem shows that the effect of a small change in a parameter of a constrained optimization problem on its maximum value can be determined by considering only the partial derivative of the objective function and the partial derivative of the constraint with respect to that parameter.

Assume the following maximization problem where the objective function *f* depends on two choice variables (x_1 and x_2) some parameter *t* (i.e., $f(x_1,x_2,t)$). Choose x_i to max the function (assume FONC and SSOC). In general, the explicit choice functions, $x_i = x_i^*(t)$, are the derived solutions to the first-order equations. If we substitute these solutions in the objective function, we obtain

$$\phi(t) = f(x_1^*(t), x_2^*(t), t)$$

Where $\phi(t)$ is the value of the objective function *f* when the x_i's that maximize *f (for a given t)* are used. This is also know at the indirect objective function.

How does $\phi(t)$ vary when *t* varies? Differentiate with respect to *t*.

$$\phi_t(t) = \frac{\partial f}{\partial x_1^*}\frac{\partial x_1^*}{\partial t} + \frac{\partial f}{\partial x_2^*}\frac{\partial x_2^*}{\partial t} + \frac{\partial f}{\partial t}$$

B/c the FOC, $\frac{\partial f}{\partial x_1^*}$ and $\frac{\partial f}{\partial x_1^*}$ equal 0, we are left with:

$$\phi_t(t) = \frac{\partial f}{\partial t}$$

This says that *t* changes, the rate of the maximum value of *f* , where x_1 and x_2 vary optimally as *t* varies, equals the rate of change of *f* as *t* varies, holding x_1 and x_2 constant.

Why do we care about this? When you begin discussing profit functions and cost functions, this is useful proving Hotelling's lemma, Shepard's lemma and Roy's identity.

Example: Profit functions maximization model

Maximize $\Pi = Pf(x_1, x_2) - w_1 x_1 - w_2 x_2$ the explicit choice functions are $x_1 = x_1^*(w_1, w_2, p)$ and $x_2 = x_2^*(w_1, w_2, p)$. If these profit maximizing levels of input are substituted into the objective function, the resulting profit level must be maximum profits attainable at those factor and output prices. In other words,

$$\Pi^*(w_1, w_2, p) = Pf(x_1^*, x_2^*) - w_1 x_1^* - w_2 x_2^*$$

The function $\Pi^*(w_1, w_2, p)$ is called the profit function or indirect profit function. Its value is always the maximum value of profits given w_1, w_2, and *p*.

How do profits vary when say w_1 changes? We could simply differentiate the objective function w.r.t. to w_1, holding not only other prices constant, but the input levels of x_1 and x_2 constant as well. In that case, we would find:

$$\frac{\partial \Pi(\cdot)}{\partial w_1} = -x_1$$

How do the <u>maximum</u> profit varies when w_1 changes? We need to differentiate the indirect profit function to get:

$$\frac{\partial \Pi^*}{\partial w_1} = p\left(\frac{\partial f_1}{\partial x_1}\frac{\partial x_1}{\partial w_1} + \frac{\partial f_1}{\partial x_2}\frac{\partial x_2}{\partial w_1}\right) - w_1 \frac{\partial x_1^*}{\partial w_1} - x_1^* - w_2 \frac{\partial x_2^*}{\partial w_1}$$

Combining the terms, yields

$$\frac{\partial \Pi^*}{\partial w_1} = \left(p\frac{\partial f}{\partial x_1} - w_1\right)\left(\frac{\partial x_1^*}{\partial w_1}\right) + \left(p\frac{\partial f}{\partial x_2} - w_2\right)\left(\frac{\partial x_2^*}{\partial w_1}\right) - x_1^*$$

The terms in parentheses on the right hand side are zero at the profit maximizing values of x_1 and x_2, thus:

$$\frac{\partial \Pi^*}{\partial w_1} = -x_1^* = \frac{\partial \Pi}{\partial w_1}$$

This says that starting at a profit-maximizing input level, the instantaneous rate of change of profit w.r.t. to a change in factor price is the same whether or not the factors are held fixed or they can vary as the factor price changes. Moreover, the value of this instantaneous rate of change is simply the negative of the factor demand function for x_1.

Equimarginal principle basically says the consumer will move along his budget line, trading x_1 for x_2 until the relative price of a unit of x_1 (price ratio) is equal to the marginal rate of substitution between x_1 and x_2. [Relate back to graph and consumer problem]

Sample Problems:

A) Let's consider a production problem. The firm wants to maximize output (y) given the following production function: $y = f(x,z)$ subject to the following expenditure constraint: $P_x*X + P_z*Z = E$. (1) Set up the optimization problem using the Lagrangian optimization technique. (2) Find the first order conditions (recall, $\frac{\partial f(x,y)}{\partial x} = MP_x$ [where MP_x is the marginal product of x]). (3) To maximize production and assuming a given level of expenditure, show the inputs (x, z) must be used in a combination such that the ratio of the MP is equal to the ratios of their prices. (4) Relate everything just calculated in (1), (2), and (3) to a graph.

B) Let's consider another production problem. The firm wants to minimize expenditures, $P_x*X + P_z*Z = E$, subject to the given production function: $y = f(x,z)$. (1) Set up the optimization problem using the Lagrangian optimization technique. (2) Find the first order conditions (recall, $\frac{\partial f(x,y)}{\partial x} = MP_x$ [where MP_x is the marginal product of x]). (3) Show that maximizing output for a given expenditure (answer calculated in (A.3 above) results in the same solution as minimizing cost for a given level of output.

In B.3, we just showed that problems can be addressed from different perspectives and identical solutions found are known as duality.

Lecture 5: Implicit Functions and Comparative Statics
References: S&B Ch 15, 22

- An *explicit* function is one in which the endogenous variable(s) are isolated from the exogenous variables, usually on the right hand side of the equation.
  - Example: $y = f(x_1, x_2, \ldots, x_n)$, or $y - f(x_1, x_2, \ldots, x_n) = 0.$
- An *implicit* function, on the other hand, does not have this property…the endogenous variable(s) cannot be separated from the exogenous variable(s).
  - Example: $G(y, x_1, x_2, \ldots, x_n) = c,\ c$ constant.
  - Specific Example: $y^5 - 5xy + 4x^2 = 0$ cannot be solved for y.
  - Economic Example:
    - The first order condition from a canonical profit maximization problem: $\max_x \ \pi(x) = pf(x) - wx,$ where p is the price of output $f(x)$ and w is the unit cost of input x.
    - FOC: $pf'(x) - w \overset{\text{set}}{=} 0.$ This is an implicit function of the form $G(x, p, w) = 0.$
    - Note that *if* this equation implicitly defines a function $x(p, w)$, then this is the input demand function for the firm….it (and its derivatives) would tell us how input demand changes with changes in output price and input price
- We'd like to know:
  1. Does $G(y, x_1, x_2, \ldots, x_n) = c,\ c$ constant, define y as a continuous function of $x_1, x_2, \ldots, x_n$ and c near any values of these variables?
  2. If so, what are the derivative properties of this function?

- IN ECONOMIC THEORY, THIS IS EXTRAORDINARILY IMPORTANT, AS FIRST ORDER CONDITIONS WHICH IMPLICITLY DEFINE THE RELATIONSHIPS BETWEEN VARIABLES AND PARAMETERS ARE UBIQUITOUS!!!!!!

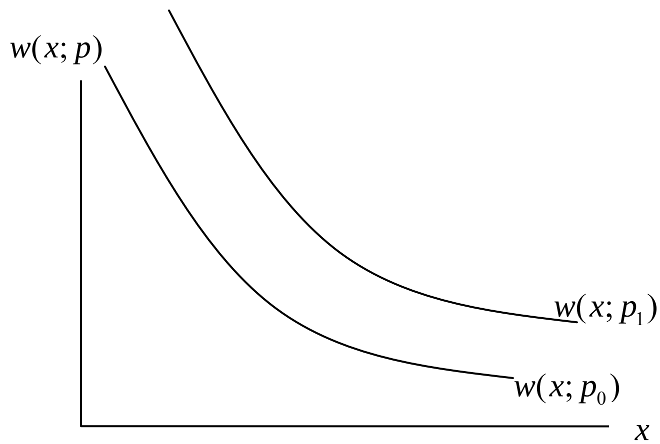
- The Implicit Function Theorem: 15.2 in Simon and Blume

  Let $G(y, x_1, x_2, \ldots, x_n)$ be a continuous function around the point $(y^*, x_1^*, x_2^*, \ldots, x_n^*)$. Suppose further that $G(y^*, x_1^*, x_2^*, \ldots, x_n^*) = c$ and that $\dfrac{\partial G(y^*, x_1^*, x_2^*, \ldots, x_n^*)}{\partial y} \neq 0.$

  THEN, there is a continuous function $y = y(x_1, x_2, \ldots, x_n)$ defined on an open ball B about $(x_1^*, x_2^*, \ldots, x_n^*)$ such that:
  1. $G(y(x_1, x_2, \ldots, x_n), x_1, x_2, \ldots, x_n) = c$ for all $(x_1, x_2, \ldots, x_n) \in B;$
  2. $y^* = y(x_1^*, x_2^*, \ldots, x_n^*);$ and

$$\text{for each } i, \ \frac{\partial y(x_1^*, x_2^*, \ldots, x_n^*)}{\partial x_i} = -\frac{\dfrac{\partial G(y^*, x_1^*, x_2^*, \ldots, x_n^*)}{\partial x_i}}{\dfrac{\partial G(y^*, x_1^*, x_2^*, \ldots, x_n^*)}{\partial y}}.$$

- Interpretation:
  - We have a continuous implicit function in the neighborhood of a set of points, and the partial derivative of this function with respect to the variable we're "solving" for (*y*) is non-zero at this set of points
  - THEN, we know that a function defining y as a function of $(x_1, x_2, \ldots, x_n)$ exists for all points "close" to our set of points that satisfies our original relationship, and we can recover how y changes with respect to all of the variables $(x_1, x_2, \ldots, x_n)$.
  - This last bit has a convenient economic interpretation: Comparative Statics!!!
- Our Economic Example:
  - $pf'(x) - w \overset{\text{set}}{=} 0$ is our implicit function. Let's assume a standard production function with diminishing marginal product.
  - First, does a function defining our input variable x as a function of output and input price, namely $x(p, w)$, exist?
    - Via IFT, it does if $\frac{\partial G(y^*, x_1^*, x_2^*, \ldots, x_n^*)}{\partial y} \neq 0$. Here, $\frac{\partial G(y, x_1, x_2, \ldots, x_n)}{\partial y} = \frac{\partial [pf'(x) - w]}{\partial x} = pf''(x) < 0$ for all x. Thus, $x(p, w)$, our input demand function, exists for all the values we care about.
  - Second, what does $x(p, w)$ look like?
    - Via IFT, $\frac{\partial x(p, w)}{\partial w} = -\frac{\dfrac{\partial [pf'(x) - w]}{\partial w}}{\dfrac{\partial [pf'(x) - w]}{\partial x}} = -\frac{(-1)}{pf''(x)} = \frac{1}{pf''(x)} < 0.$
    - So What? $\frac{\partial x(p, w)}{\partial w}$ tells us how the quantity demanded for input x changes with the price of x. We just found a negative relationship between the two. This, then, is the LAW OF DEMAND!!!
    - Note we could also define the inverse demand…



Input Demand from a One-Input
Production Problem, $p_1 > p_0$

- Via IFT, $\dfrac{\partial x(p,w)}{\partial p} = -\dfrac{\dfrac{\partial\left[pf'(x)-w\right]}{\partial p}}{\dfrac{\partial\left[pf'(x)-w\right]}{\partial x}} = -\dfrac{f'(x)}{pf''(x)} > 0.$
- So What? $\dfrac{\partial x(p,w)}{\partial p}$ tells us how the quantity demanded for input x changes with the price of output. We just found a positive relationship between the two. This means that if output price goes up, *all else equal* (ceteris paribus), demand for input x increases.
- Graphically, this is a shift in the input demand curve

- The Implicit Function Theorem can be generalized to systems of implicit functions, typical in a multivariate framework. The main difference is that the conditions under which a system of functions exists depends on the determinant of the matrix of first derivatives, and the partial derivatives of each endogenous variable with respect to each exogenous variable.
- To illustrate, let's use the canonical consumer problem:

$$\max_{x_1,x_2}\ u(x_1,x_2)$$
$$s.t.\ p_1x_1 + p_2x_2 \le I,$$

where we assume a concave utility function.

Defining the Lagrangian $L = u(x_1, x_2) - \lambda[p_1 x_1 + p_2 x_2 - I]$, the first-order conditions are:

$$\frac{\partial L}{\partial x_1} = u_{x_1}(x_1, x_2) - \lambda p_1 \overset{\text{set}}{=} 0,$$

$$\frac{\partial L}{\partial x_2} = u_{x_2}(x_1, x_2) - \lambda p_2 \overset{\text{set}}{=} 0,$$

$$\frac{\partial L}{\partial \lambda} = p_1 x_1 + p_2 x_2 - I \overset{\text{set}}{=} 0.$$

Combining the first two FOCs, we obtain

$$G_1(x_1, x_2, p_1, p_2) = \frac{u_{x_1}(x_1, x_2)}{p_1} - \frac{u_{x_2}(x_1, x_2)}{p_2} = 0$$

$$G_2(x_1, x_2, p_1, p_2, I) = p_1 x_1 + p_2 x_2 - I = 0.$$

- Note this is two equations in two unknowns…we'd like to define optimal x_1 and x_2 as functions of the known parameters of the problem (i.e., demand functions).
- We'd like to know how the demand curves $x_1(p_1, p_2, I)$ and $x_2(p_2, p_1, I)$, assuming they exist, respond to, say, changes in p_1.
    - Posit the existence of $x_1(p_1, p_2, I)$ and $x_2(p_2, p_1, I)$, and substitute into the combined FOCs:

$$G_1(x_1(p_1, p_2, I), x_2(p_2, p_1, I), p_1, p_2)$$
$$= \frac{u_{x_1}(x_1(p_1, p_2, I), x_2(p_2, p_1, I))}{p_1} - \frac{u_{x_2}(x_1(p_1, p_2, I), x_2(p_2, p_1, I))}{p_2} = 0$$
$$G_2(x_1(p_1, p_2, I), x_2(p_2, p_1, I), p_1, p_2, I)$$
$$= p_1 x_1(p_1, p_2, I) + p_2 x_2(p_2, p_1, I) - I = 0$$

    - Take the partial derivative of each equation w.r.t. p_1, which we can do since these are identities:

$$\frac{\partial G_1(\cdot)}{\partial x_1}\frac{\partial x_1(\cdot)}{\partial p_1} + \frac{\partial G_1(\cdot)}{\partial x_2}\frac{\partial x_2(\cdot)}{\partial p_1} + \frac{\partial G_1(\cdot)}{\partial p_1} = 0$$
$$\frac{\partial G_2(\cdot)}{\partial x_1}\frac{\partial x_1(\cdot)}{\partial p_1} + \frac{\partial G_2(\cdot)}{\partial x_2}\frac{\partial x_2(\cdot)}{\partial p_1} + \frac{\partial G_2(\cdot)}{\partial p_1} = 0$$

    Set this up in matrix form:

$$\begin{bmatrix} \frac{\partial G_1(\cdot)}{\partial x_1} & \frac{\partial G_1(\cdot)}{\partial x_2} \\ \frac{\partial G_2(\cdot)}{\partial x_1} & \frac{\partial G_2(\cdot)}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1(\cdot)}{\partial p_1} \\ \frac{\partial x_2(\cdot)}{\partial p_1} \end{bmatrix} = \begin{bmatrix} -\frac{\partial G_1(\cdot)}{\partial p_1} \\ -\frac{\partial G_2(\cdot)}{\partial p_1} \end{bmatrix}$$

This is in the form AX=b in matrix notation, so a solution exists so long as the inverse of the matrix A exists, which it does so long as its determinant does not equal zero. This is the condition for the systems version of IFT.

This is true so long as the utility function is quasi-concave [proof on your own].

- In order to solve for the partial derivatives, you can use any means you like, but Cramers' rule comes in pretty handy (Theroem 9.4b in S&B, p. 194):

  The unique solution $\mathbf{x} = (x_1, \ldots, x_n)$ of the *n* x *n* system $\mathbf{Ax} = \mathbf{b}$ is:

  $$x_i = \frac{\det\ \mathbf{B}_i}{\det\ \mathbf{A}},\ i = 1, \ldots, n,$$

  where $\mathbf{B}_i$ is the matrix $\mathbf{A}$ with the right-hand side $\mathbf{b}$ replacing the *i*th column of $\mathbf{A}$.

  - For our example above, this implies:

    $$\det\ \mathbf{B}_1 = \begin{vmatrix} -\frac{\partial G_1(\cdot)}{\partial p_1} & \frac{\partial G_1(\cdot)}{\partial x_2} \\ \frac{\partial G_2(\cdot)}{\partial x_1} & \frac{\partial G_2(\cdot)}{\partial x_2} \end{vmatrix},\ \det\ \mathbf{A} = \begin{vmatrix} \frac{\partial G_1(\cdot)}{\partial x_1} & \frac{\partial G_1(\cdot)}{\partial x_2} \\ \frac{\partial G_2(\cdot)}{\partial x_1} & \frac{\partial G_2(\cdot)}{\partial x_2} \end{vmatrix}, \text{and}$$

    $$\frac{\partial x_1(\cdot)}{\partial p_1} = \det \mathbf{B}_1 / \det \mathbf{A}.$$

- Note also that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

PRACTICE PROBLEMS

1. For all of the questions below, assume the following: you are given a $C^{(2)}$ function $f(x, y; \boldsymbol{\alpha}) : \Re^2 \to \Re_+$, where $\boldsymbol{\alpha} \overset{\text{def}}{=} (\gamma, \delta) \in \Re^2_{++}$ is a parameter vector, with $f_x(x, y; \boldsymbol{\alpha}) > 0,\ f_y(x, y; \boldsymbol{\alpha}) < 0,\ \frac{\partial f(x, y; \boldsymbol{\alpha})}{\partial \gamma} > 0,$ and $\frac{\partial f(x, y; \boldsymbol{\alpha})}{\partial \delta} < 0.$
   a. In laymen's terms, interpret these assumptions.
   b. Suppose $\gamma f(x, y; \boldsymbol{\alpha}) - \delta y = 0.$

i. Does a function $y = y(x;\boldsymbol{\alpha})$ exist? How do you know?
ii. Is this function continuous? Explain.
iii. Find an expression for $y'(x;\boldsymbol{\alpha})$ and graph $y(x;\boldsymbol{\alpha})$ in Y-X space (y on the vertical axis).
iv. Now assume $x = \bar{x} > 0,$ and $\gamma f(\bar{x}, y;\boldsymbol{\alpha}) - \delta y = 0.$ Find the comparative statics of $y(\bar{x};\boldsymbol{\alpha})$ with respect to each of the elements of $\boldsymbol{\alpha},$ and sign them if possible. Interpret your results.

c. Now assume that x is no longer fixed, and you are given the following system of equations that potentially defines an equilibrium (if it exists) for x and y:

$$\gamma f(x, y;\boldsymbol{\alpha}) - \delta y = 0$$
$$x + \gamma y = 0.$$

i. Is the solution $\left(x^*(\boldsymbol{\alpha}), y^*(\boldsymbol{\alpha})\right)$ for the above system well-defined? How do you know?
ii. Are $x^*(\boldsymbol{\alpha})$ and $y^*(\boldsymbol{\alpha})$ continuous? Of what degree? Explain.
iii. Find the comparative statics of $\left(x^*(\boldsymbol{\alpha}), y^*(\boldsymbol{\alpha})\right)$ with respect to each of the elements of $\boldsymbol{\alpha},$ and sign them if possible. Interpret your results. (HINT: First substitute the optimal solution into the system given above, then totally differentiate with respect to the parameter of interest and use Cramer's rule to solve).

2. Consider a utility function $U(x, y)$.
   a. Using the implicit function theorem, show that indifference curves are negatively sloped so long as marginal utilities with respect to goods x and y are positive.
   b. Does this condition ensure a diminishing marginal rate of substitution (DMRS)?
   c. Assuming a DMRS, graph a few indifference curves.

3. Consider a monopolist facing an inverse demand curve for a normal good $p(q; y)$, where q is quantity sold and y is per-capita income. Total cost for this monopolist is given by $c(q;\gamma)$, where γ is a measure of technology such that $\gamma_i > \gamma_j$ implies $c'(q;\gamma_i) < c'(q;\gamma_j)$.
   a. Using calculus, sign, interpret, and graph the properties of the monopolist's marginal revenue and marginal cost curves.
   b. Set up the monopolist's profit maximization problem and find the first and second order conditions.
   c. Using the implicit function theorem, qualitatively describe how the monopolist's output changes with per capita income and technology. Verify your answer graphically.

Lecture 6: Matrix Algebra in Economics
References: S&B Ch 8, 15

What is a matrix? A matrix is a rectangular array of numbers. The numerical content of a table of data is indeed a perfect input for a matrix.

- Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 6 & 4 \\ 13 & 8 & 5 \\ 8 & 9 & 5 \end{pmatrix}$
- A matrix with *n* rows and *k* columns is called a *k* x *n* ("k by n" is the *dimension*) matrix. Thus, **A** is a 3x3 matrix. A matrix with the same number of rows and columns, like A, is called a *square matrix*.
- The number in the i_{th} row and j_{th} column of a matrix is referred to as an *element*, or *entry* of the matrix, and is often represented with a_{ij}. In matrix A, $a_{23} = 5$
- A *vector* is a matrix with many rows and only one column (*column vector*), or one row and many columns (*row vector*).

Why we bother learning matrix algebra? The use of matrices is pervasive in advanced economics and econometrics. Matrix algebra

- provides a compact way of representing systems of equations, no matter how large and complicated.
- allows testing for the existence of a solution to a system of equations, and solve for it.
- The catch is that matrix algebra is applicable to *linear*-equation systems. When the equations are nonlinear, matrix algebra can be used to provide linear approximations that hold locally (i.e. in a small neighborhood about a point)

Examples of applications.

- **Macroeconomics**: Consider the Keynesian national-income macroeconomic model
  - $Y = C + I_0 + G_0$ (equilibrium condition)
  - $C = a + bY$ (behavioral consumption function)

  Where Y and C are national income and consumption expenditure variables (endogenous, or unknown and determined by the model), and I_0 and G_0 are investment and government expenditure (exogenous or predetermined). a>0 and 0<b<1 are model parameters (autonomous consumption and marginal propensity to consume).
- We will show that the system can be represented with the simple notation **Ax = d** , and we can find the values of Y and C satisfying both equations (solve) for any given set of parameters and exogenous variable. Such solution can be calculated using matrix algebra as $\bar{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{d}$
- **Econometrics.** You have seen in lecture 3 that parameter estimates for the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ can be obtained via OLS, which consists in solving a minimization problem (get the 2 normal equations $\frac{\sum_{i=1}^{N} e_i}{\partial \beta_0} \overset{set}{=} 0; \frac{\sum_{i=1}^{N} e_i}{\partial \beta_1} \overset{set}{=} 0$, solve for β_0 and β_1).

Consider now the model with k explanatory variables
$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$. The OLS estimators for each parameter can be obtained by solving the *k* normal equations:

$$\sum_{i=1}^{N} y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{N} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik}$$

$$\sum_{i=1}^{N} y_i x_{i2} = \hat{\beta}_1 \sum_{i=1}^{N} x_{i2} + \hat{\beta}_2 \sum_{i=1}^{N} x^2{}_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik} x_{t2}$$

.

.

.

$$\sum_{i=1}^{N} y_i x_{ik} = \hat{\beta}_1 \sum_{i=1}^{N} x_{ik} + \hat{\beta}_2 \sum_{i=1}^{N} x_{i2} x_{ik} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik}^2$$

This is a linear system of *k* equations in *k* unknowns. It is also quite revolting even to just look at. We will show that the system can be represented in matrix notation as $\mathbf{X'Y} = \mathbf{X'X}\boldsymbol{\beta}$ and the *k* parameters solving the system of equations are found as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'Y})$$

Rules of Matrix algebra

- **Addition and subtraction**
  Conformability rule: two matrices can be added (subtracted) if and only if they have the same dimension. When this condition is met, the matrices are said to be *conformable for addition (subtraction)*. More in detail, if we consider the *2x2* matrices **A** and **B:**

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

  To give a numerical example:

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 4 & 9 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{pmatrix}$$

  Subtraction works analogously.
- **Scalar multiplication**
  Any matrix can be multiplied by an ordinary number (scalar). Each element of the original matrix is multiplied by the scalar. Thus, for the *2x2* matrix **A** and the scalar r:

$$r\mathbf{A} = r\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} ra_{11} & ra_{12} \\ ra_{21} & ra_{22} \end{pmatrix}$$

- **Matrix multiplication**
  - **Conformability rule**: two matrices **A** and **B** can be multiplied if only if the "Number of columns of **A**= Number of rows of B". For the **AB**=**C** matrix to exist, **A** must be *k* x *m* and **B** must be *m* x *n*.
  - The resulting dimension of **C** is *k* x *n*.

- The $(i,j)_{th}$ entry of the matrix **C** is defined to be $\sum_{h=1}^{m} a_{ih} b_{hj}$. That is, to obtain the $(i,j)_{th}$ entry of C, multiply the i_{th} row of A and the j_{th} column of B as follows:
- $\begin{pmatrix} a_{i1} & a_{i2} & \cdots a_{im} \end{pmatrix} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{pmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}$. This is referred to as the **dot product** of the two vectors.
- Notice that premultiplication and postmultiplication are not the same thing for matrices. That is, in most cases, $AB \neq BA$. The commutative law of multiplication does not apply to matrices.
- Numerical example:

$$\underset{3x3}{A} = \begin{pmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{pmatrix};\ \underset{3x2}{B} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix};\ \underset{3x2}{C} = \begin{pmatrix} 3-2+6 & 12-5+12 \\ 1+0+9 & 4+0+18 \\ 4+0+6 & 16+0+12 \end{pmatrix} = \begin{pmatrix} 7 & 19 \\ 10 & 22 \\ 10 & 28 \end{pmatrix}$$

- We can now easily show that the Keynesian macroeconomic model we introduced earlier is indeed represented by $\mathbf{Ax = d}$, if we define
  - $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -b & 1 \end{pmatrix}$ a matrix of coefficients
  - $\mathbf{x} = \begin{pmatrix} Y \\ C \end{pmatrix}$ a vector of endogenous variables and
  - $\mathbf{d} = \begin{pmatrix} I_0 + G_0 \\ a \end{pmatrix}$ a vector of exogenous variables.
  - $\mathbf{Ax = d} \equiv \begin{pmatrix} 1 & -1 \\ -b & 1 \end{pmatrix} \begin{pmatrix} Y \\ C \end{pmatrix} = \begin{pmatrix} I_0 + G_0 \\ a \end{pmatrix} \equiv \begin{matrix} Y - C = I_0 + G_0 \\ -bY + c = a \end{matrix}$

- **Transpose**
  - The *transpose* of a matrix **A**, indicated as $\mathbf{A}' \equiv \mathbf{A}^{\mathbf{T}}$, is obtained by interchanging the rows and column of a matrix.
    - The transpose of a row vector is a column vector, and vice-versa
    - Dimension: if A is *n x k*, A' is *k x n*
    - For $\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix};\ \mathbf{A'} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$

- For $\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; $\mathbf{I'} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This particular matrix (with ones in the NW-SE diagonal, and zeros in all other entries) is called the *identity* matrix, and has a series of peculiarities. You can think about it as the matrix-equivalent of the scalar 1.

- **Determinant of a matrix**
  - Notation: $|\mathbf{A}|$ indicates the *determinant* of the matrix A. Determinants are defined only for square matrices.
  - If $|\mathbf{A}| = \mathbf{0}$, A is said to be *singular*. If $|\mathbf{A}| \neq 0$, we say **A** is *non-singular*. This implies that all the columns of matrix **A** are *linearly independent*: none of the columns of **A** is a linear function of another column of **A** (more later).
  - The determinant of the *2x2* square matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, is computed as $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$. A simple rule exists also for *3x3* matrices. For bigger matrices, use a PC.
  - $|\mathbf{A}| = \begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix} = 1 - b$

Inverse

- What about matrix division? Divisions, in matrix algebra do not exist (why?). The *inverse* of a matrix is somehow related to it.
- Notation: $\mathbf{A}^{-1}$ indicates the inverse of the matrix A. Such inverse may or or may not exist. If it exists, it is unique.
- Definition: $\mathbf{A}^{-1}$ is such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ (pre- or post-multiplication of a matrix by its inverse returns the identity matrix)
- Conditions for the existence of A^{-1}
  - **A** is a square matrix
  - $|\mathbf{A}| \neq 0$. i.e. **A** is nonsingular.
  - Inverses can be used to compute the solution to a linear system
  - Calculating inverses: finding the inverses of a large matrix can be quite complex and annoying. Luckily, using modern computers take a fraction of a second. For the time being, it will suffice to know that, when A is a diagonal matrix as:

$$A=\begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & a_{nn} \end{pmatrix}\text{; then } A^{-1}=\begin{pmatrix} \frac{1}{a_{11}} & 0 & 0 & 0 \\ 0 & \frac{1}{a_{22}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{a_{nn}} \end{pmatrix}$$

- Going back to the macroeconomic model, your computer tells you that $\mathbf{A}^{-1}=\frac{1}{1-b}\begin{pmatrix} 1 & 1 \\ b & 1 \end{pmatrix}$. Using $\overline{\mathbf{x}}=\mathbf{A}^{-1}\mathbf{d}$ we have

  $$\begin{pmatrix} \overline{Y} \\ \overline{C} \end{pmatrix}=\frac{1}{1-b}\begin{pmatrix} 1 & 1 \\ b & 1 \end{pmatrix}\begin{pmatrix} I_0+G_0 \\ a \end{pmatrix}=\frac{1}{1-b}\begin{pmatrix} I_0+G_0+a \\ b\left(I_0+G_0\right)+a \end{pmatrix}$$

Solving a system of linear equations: Cramer's rule

- Cramer's rule gives a handy tool to find solutions to systems of linear equations without looking for an inverse
- Let **A** be a nonsingular matrix and $\mathbf{Ax}=\mathbf{d}$ represent a system of linear equations. Each element in the vector ***x*** can be computed as

  $x_i=\frac{|\mathbf{D_i}|}{|\mathbf{A}|}$; where $\mathbf{D_i}$ is the matrix **A** with **d** replacing the i_{th} column
- Using our macro model:

  $$\overline{Y}=\frac{\begin{vmatrix} I_0+G_0 & -1 \\ a & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}}=\frac{I_0+G_0+a}{1-b}.$$ $\overline{C}$ can be derived analogously.

Back to the least squares problem: practice problem

Consider $\mathbf{X}=\begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{N1} \end{pmatrix}$ and $\mathbf{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$. You should be able now to show that indeed the matrix representation $\hat{\boldsymbol{\beta}}=\left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'Y}\right)$ of the Ordinary Least Square estimator is equivalent to

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \overline{y} - \hat{\beta}_1 \overline{x} \\ \dfrac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^{\,2} - \left(\sum_{i=1}^{N} x_i\right)^2} = \dfrac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \end{pmatrix}$$

Additional practice problems

1. Solve for $\overline{C}$ in the national income model using Kramer rule.
2. Write a (large) system of linear equations in matrix form. I'll see if there is any suggestion coming from the other sets of notes

Consider the vector $\hat{\boldsymbol{\beta}} = \left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'Y}\right)$; where $\underset{\mathbf{nxk}}{\mathbf{X}}$ and $\underset{\mathbf{nx1}}{\mathbf{Y}}$.

1. Must X be a square matrix? Must $X'X$? Must $X'Y$?
2. What is the dimension of $\hat{\boldsymbol{\beta}}$? How many unknowns are implied by the above system?

Lecture 6: Matrix Algebra in Economics
References: S&B Ch 8, 15

What is a matrix? A matrix is a rectangular array of numbers. The numerical content of a table of data is indeed a perfect input for a matrix.

- Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 6 & 4 \\ 13 & 8 & 5 \\ 8 & 9 & 5 \end{pmatrix}$
- A matrix with *n* rows and *k* columns is called a *k* x *n* ("k by n" is the *dimension*) matrix. Thus, **A** is a 3x3 matrix. A matrix with the same number of rows and columns, like A, is called a *square matrix*.
- The number in the i_{th} row and j_{th} column of a matrix is referred to as an *element*, or *entry* of the matrix, and is often represented with a_{ij}. In matrix A, $a_{23} = 5$
- A *vector* is a matrix with many rows and only one column (*column vector*), or one row and many columns (*row vector*).

Why we bother learning matrix algebra? The use of matrices is pervasive in advanced economics and econometrics. Matrix algebra

- provides a compact way of representing systems of equations, no matter how large and complicated.
- allows testing for the existence of a solution to a system of equations, and solve for it.
- The catch is that matrix algebra is applicable to *linear*-equation systems. When the equations are nonlinear, matrix algebra can be used to provide linear approximations that hold locally (i.e. in a small neighborhood about a point)

Examples of applications.

- **Macroeconomics**: Consider the Keynesian national-income macroeconomic model
  - o $Y = C + I_0 + G_0$ (equilibrium condition)
  - o $C = a + bY$ (behavioral consumption function)

  Where Y and C are national income and consumption expenditure variables (endogenous, or unknown and determined by the model), and I_0 and G_0 are investment and government expenditure (exogenous or predetermined). a>0 and

$0<b<1$ are model parameters (autonomous consumption and marginal propensity to consume).

- We will show that the system can be represented with the simple notation **Ax = d** , and we can find the values of Y and C satisfying both equations (solve) for any given set of parameters and exogenous variable. Such solution can be calculated using matrix algebra as $\bar{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{d}$
- **Econometrics.** You have seen in lecture 3 that parameter estimates for the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ can be obtained via OLS, which consists in solving a minimization problem (get the 2 normal equations $\frac{\sum_{i=1}^{N} e_i}{\partial \beta_0} \overset{set}{=} 0; \frac{\sum_{i=1}^{N} e_i}{\partial \beta_1} \overset{set}{=} 0$, solve for β_0 and β_1).

  Consider now the model with k explanatory variables $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$. The OLS estimators for each parameter can be obtained by solving the *k* normal equations:

$$\sum_{i=1}^{N} y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{N} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik}$$

$$\sum_{i=1}^{N} y_i x_{i2} = \hat{\beta}_1 \sum_{i=1}^{N} x_{i2} + \hat{\beta}_2 \sum_{i=1}^{N} x^2_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik} x_{t2}$$

$$\vdots$$

$$\sum_{i=1}^{N} y_i x_{ik} = \hat{\beta}_1 \sum_{i=1}^{N} x_{ik} + \hat{\beta}_2 \sum_{i=1}^{N} x_{i2} x_{ik} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} x_{ik}^2$$

  This is a linear system of *k* equations in *k* unknowns. It is also quite revolting even to just look at. We will show that the system can be represented in matrix notation as $\mathbf{X'Y} = \mathbf{X'X}\boldsymbol{\beta}$ and the *k* parameters solving the system of equations are found as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'Y})$$

Rules of Matrix algebra

- **Addition and subtraction**

  Conformability rule: two matrices can be added (subtracted) if and only if they have the same dimension. When this condition is met, the matrices are said to be *conformable for addition (subtraction)*. More in detail, if we consider the *2x2* matrices **A** and **B:**

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix}$$

  To give a numerical example:

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} 4 & 9 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{pmatrix}$$

  Subtraction works analogously.

- **Scalar multiplication**

Any matrix can be multiplied by an ordinary number (scalar). Each element of the original matrix is multiplied by the scalar. Thus, for the *2x2* matrix **A** and the scalar r:

$$r\mathbf{A} = r\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} ra_{11} & ra_{12} \\ ra_{21} & ra_{22} \end{pmatrix}$$

- o **Matrix multiplication**
  - **Conformability rule**: two matrices **A** and **B** can be multiplied if only if the "Number of columns of **A**= Number of rows of B". For the **AB**=**C** matrix to exist, **A** must be *k* x *m* and **B** must be *m* x *n*.
  - The resulting dimension of **C** is *k* x *n*.
  - The $(i,j)_{th}$ entry of the matrix **C** is defined to be $\sum_{h=1}^{m} a_{ih}b_{hj}$. That is, to obtain the $(i,j)_{th}$ entry of C, multiply the i_{th} row of A and the j_{th} column of B as follows:
  - $\begin{pmatrix} a_{i1} & a_{i2} & \cdots a_{im} \end{pmatrix}\begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{pmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}$. This is referred to as the **dot product** of the two vectors.
  - Notice that premultiplication and postmultiplication are not the same thing for matrices. That is, in most cases, $AB \neq BA$. The commutative law of multiplication does not apply to matrices.
  - Numerical example:

$$\underset{3x3}{A} = \begin{pmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{pmatrix};\ \underset{3x2}{B} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix};\ \underset{3x2}{C} = \begin{pmatrix} 3-2+6 & 12-5+12 \\ 1+0+9 & 4+0+18 \\ 4+0+6 & 16+0+12 \end{pmatrix} = \begin{pmatrix} 7 & 19 \\ 10 & 22 \\ 10 & 28 \end{pmatrix}$$

  - o We can now easily show that the Keynesian macroeconomic model we introduced earlier is indeed represented by $\mathbf{Ax = d}$, if we define
    - o $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -b & 1 \end{pmatrix}$ a matrix of coefficients
    - o $\mathbf{x} = \begin{pmatrix} Y \\ C \end{pmatrix}$ a vector of endogenous variables and
    - o $\mathbf{d} = \begin{pmatrix} I_0 + G_0 \\ a \end{pmatrix}$ a vector of exogenous variables.
    - o $\mathbf{Ax = d} \equiv \begin{pmatrix} 1 & -1 \\ -b & 1 \end{pmatrix}\begin{pmatrix} Y \\ C \end{pmatrix} = \begin{pmatrix} I_0 + G_0 \\ a \end{pmatrix} \equiv \begin{matrix} Y - C = I_0 + G_0 \\ -bY + c = a \end{matrix}$

- **Transpose**
  - The *transpose* of a matrix **A**, indicated as $\mathbf{A}' \equiv \mathbf{A}^{\mathrm{T}}$, is obtained by interchanging the rows and column of a matrix.
    - o The transpose of a row vector is a column vector, and vice-versa

- Dimension: if A is *n x k*, $A^{'}$ is *k x n*
- For $\mathbf{A}=\begin{pmatrix}1 & 4\\ 2 & 5\\ 3 & 6\end{pmatrix}$; $\mathbf{A'}=\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}$
- For $\mathbf{I}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$; $\mathbf{I'}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$. This particular matrix (with ones in the NW-SE diagonal, and zeros in all other entries) is called the *identity* matrix, and has a series of peculiarities. You can think about it as the matrix-equivalent of the scalar 1.

- **Determinant of a matrix**
  - Notation: $|\mathbf{A}|$ indicates the *determinant* of the matrix A. Determinants are defined only for square matrices.
  - If $|\mathbf{A}|=\mathbf{0}$, A is said to be *singular*. If $|\mathbf{A}|\neq 0$, we say **A** is *non-singular*. This implies that all the columns of matrix **A** are *linearly independent*: none of the columns of **A** is a linear function of another column of **A** (more later).
  - The determinant of the *2x2* square matrix $\mathbf{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$, is computed as $|\mathbf{A}|=a_{11}a_{22}-a_{12}a_{21}$. A simple rule exists also for *3x3* matrices. For bigger matrices, use a PC.
  - $|\mathbf{A}|=\begin{vmatrix}1 & -1\\ -b & 1\end{vmatrix}=1-b$

Inverse

- What about matrix division? Divisions, in matrix algebra do not exist (why?). The *inverse* of a matrix is somehow related to it.
- Notation: $\mathbf{A}^{-1}$ indicates the inverse of the matrix A. Such inverse may or or may not exist. If it exists, it is unique.
- Definition: $\mathbf{A}^{-1}$ is such that $\mathbf{AA^{-1}=A^{-1}A=I}$ (pre- or post-multiplication of a matrix by its inverse returns the identity matrix)
- Conditions for the existence of A^{-1}
  - **A** is a square matrix
  - $|\mathbf{A}|\neq 0$. i.e. **A** is nonsingular.
  - Inverses can be used to compute the solution to a linear system
  - Calculating inverses: finding the inverses of a large matrix can be quite complex and annoying. Luckily, using modern computers take a fraction of a second. For the time being, it will suffice to know that, when A is a diagonal matrix as:

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & a_{nn} \end{pmatrix}; \text{ then } A^{-1} = \begin{pmatrix} \frac{1}{a_{11}} & 0 & 0 & 0 \\ 0 & \frac{1}{a_{22}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{a_{nn}} \end{pmatrix}$$

- Going back to the macroeconomic model, your computer tells you that $\mathbf{A}^{-1} = \frac{1}{1-b}\begin{pmatrix} 1 & 1 \\ b & 1 \end{pmatrix}$. Using $\overline{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{d}$ we have

  $$\begin{pmatrix} \overline{Y} \\ \overline{C} \end{pmatrix} = \frac{1}{1-b}\begin{pmatrix} 1 & 1 \\ b & 1 \end{pmatrix}\begin{pmatrix} I_0 + G_0 \\ a \end{pmatrix} = \frac{1}{1-b}\begin{pmatrix} I_0 + G_0 + a \\ b\left(I_0 + G_0\right) + a \end{pmatrix}$$

Solving a system of linear equations: Cramer's rule

- Cramer's rule gives a handy tool to find solutions to systems of linear equations without looking for an inverse
- Let **A** be a nonsingular matrix and $\mathbf{Ax} = \mathbf{d}$ represent a system of linear equations. Each element in the vector ***x*** can be computed as

  $x_i = \frac{|\mathbf{D_i}|}{|\mathbf{A}|}$; where $\mathbf{D_i}$ is the matrix **A** with **d** replacing the i_{th} column
- Using our macro model:

  $\overline{Y} = \frac{\begin{vmatrix} I_0 + G_0 & -1 \\ a & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{I_0 + G_0 + a}{1-b}$. $\overline{C}$ can be derived analogously.

Back to the least squares problem: practice problem

Consider $\mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{N1} \end{pmatrix}$ and $\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$. You should be able now to show that indeed the matrix representation $\hat{\boldsymbol{\beta}} = \left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'Y}\right)$ of the Ordinary Least Square estimator is equivalent to

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \dfrac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^{\,2} - \left(\sum_{i=1}^{N} x_i\right)^2} = \dfrac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2} \end{pmatrix}$$

Additional practice problems

3. Solve for $\bar{C}$ in the national income model using Cramer rule.
4. Use Cramer rule to solve the following equation systems:
   a. $$\begin{aligned} 3x_1 - 2x_2 &= 11 \\ 2x_1 + x_2 &= 12 \end{aligned}$$
   b. $$\begin{aligned} 6x_1 + 9x_2 &= 15 \\ 7x_1 - 3x_2 &= 3 \end{aligned}$$

Consider the vector $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'Y})$; where $\underset{\mathbf{nxk}}{\mathbf{X}}$ and $\underset{\mathbf{nx1}}{\mathbf{Y}}$.

3. Must X be a square matrix? Must $X'X$? Must $X'Y$?
4. What is the dimension of $\hat{\boldsymbol{\beta}}$? How many unknowns are implied by the above system?
5. Consider a simple linear model in which the investment of the ith consumer in stocks is a linear function of a constant term, income, his age and a stochastic error term. You collected a sample on US consumers, and want to estimate the relationship

$$inv_i = \beta_0 + \beta_1 income_i + \beta_2 age_i + \varepsilon_i \qquad i = 1,\ldots,N$$

(betas are parameters to be estimated, N represents the total number of observations collected)

   a. Specify the model in the matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. What are the elements of each vector, and matrix? Note that the presence of a constant implies a column of ones in the X matrix

Now imagine we collect another sample from Italian investors. You might want to find out if US and Italian consumers' investment decisions are affected in the same way by income and age. In other words, you want to estimate the models

$$1) inv_i = \beta_0^{US} + \beta_1^{US} income_i + \beta_2^{US} age_i + \varepsilon_i \qquad i = 1,\ldots,N^{US}$$
$$2) inv_i = \beta_0^{It} + \beta_1^{It} income_i + \beta_2^{It} age_i + \varepsilon_i \qquad i = 1,\ldots,N^{It}$$

   b. Specify the model in the matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. You should be able to define only one X matrix, one Y, one $\boldsymbol{\beta}$ and one error vector.
   Hint: stack the vectors implied by equations 1) and 2). For the matrix X, use zeros where needed so that the model in 1) and 2) is obtained.

   c. Now, imagine you estimated the parameters of the model in b). We want to check if indeed Us and Italian investors react differently to variations in income and age. What are the coordinates of each pair of relevant parameter comparisons in the vector $\boldsymbol{\beta}$?