Mathematical Prerequisites for
Econ. 7005, "Microeconomic Theory I," and
Econ. 7007, "Macroeconomic Theory I,"
at the University of Utah

by Gabriel A. Lozada

August 2023

The contents of Sections 1–6 below are required for basic consumer and producer theory, which is usually taught at the very beginning of Econ. 7005. The contents of Section 7 ("§7") are not required until the topic of uncertainty is reached, which is typically not until Econ. 7006. Sections 1–4 ("§§1–4") are a shortened version of a more complete treatment described in lecture notes I wrote for Econ. 7001; those notes are available on the web site.

Among the terms and notation which I do not explain but which you will need to know are the following.

General background:

- letters of the Greek alphabet commonly used in mathematics (see Table 1);
- a "0" subscript, "naught";
- x-prime x' and x-hat $\hat{x}$ and x-tilde $\tilde{x}$;
- strict inequality, weak inequality;
- the difference between "$f(x) = 2$" and "$f(x) \equiv 2$";
- functional composition, $f(g(x)) = (f \circ g)(x)$;
- $x \in A$ is equivalent to $A \ni x$ ("A owns x");
- $A \subset B$, $A \subseteq B$, $A \cup B$, $A \cap B$;
- set difference, $A \setminus B$ ("set minus");
- the complement of a set; if Ω is the "universal set" or the "universe"and if $A \subseteq \Omega$, then the complement of A is written $\Omega \setminus A$ or A^C or $\complement_\Omega A$ or $\complement A$ or $\bar{A}$;
- $A \times B$, the "Cartesian product" of two sets;
- R^n, R^{n+}, R^{n++} or $\mathbf{R}^n$, $\mathbf{R}^{n+}$, $\mathbf{R}^{n++}$ or $\Re^n$, $\Re^{n+}$, $\Re^{n++}$ or $\mathbb{R}^n$, $\mathbb{R}^{n+}$, $\mathbb{R}^{n++}$ or $\mathbf{R}^n_+$, $\mathbf{R}^n_{++}$ etc.;
- open interval (of the real line $\mathbf{R}^1$), closed interval, half-open interval (which is the same as a half-closed interval); notations $[a, b]$, (a, b), $[a, b)$, $(a, b]$, or the nonstandard $[a, b[$ and $]a, b]$;

α		alpha
β		beta
γ	Γ	gamma
δ	Δ	delta
ϵ, ε		epsilon
ζ		zeta
η		eta
θ, ϑ	Θ	theta
ι		iota
κ		kappa
λ	Λ	lambda
μ		mu
ν		nu
ξ	Ξ	xi (pronounced 'zi')
o		omicron (same as Roman 'o')
π	Π	pi
ρ, ϱ		rho (the 'h' is silent)
σ	Σ	sigma
τ		tau
υ	Υ	upsilon
ϕ, φ	Φ	phi
χ		chi (pronounced 'ki')
ψ	Ψ	psi (the 'p' is silent)
ω	Ω	omega

Table 1. Upper-case Greek letters are not listed if they are the same as the Roman upper-case form. Neither the lower-case omicron nor the lower-case upsilon is ever used in mathematical formulas; neither are variant forms of the lower-case pi ("ϖ") and sigma ("ς"). The lower-case epsilon ("ϵ") should not be confused with the set-inclusion sign ("$\in$").

- open subset of $\mathbf{R}^n$ and closed subset of $\mathbf{R}^n$;
- the following potentially confusing notations:
  - a set having two elements $\{a, b\}$
  - the two-dimensional vector $(a, b) \in \mathbf{R}^2$
  - the open interval $(a, b) \in \mathbf{R}^1$
  - the closed interval $[a, b] \in \mathbf{R}^1$.
- bounded set;
- vector inequalities $\mathbf{x} \geq \mathbf{y}$, $\mathbf{x} > \mathbf{y}$, $\mathbf{x} \gg \mathbf{y}$, $\mathbf{x} \geqq \mathbf{y}$, $\mathbf{x} \ggg \mathbf{y}$ (usage varies with authors; see p. 475 of Varian);
- "scalar product," also known as "dot product";
- monotonic function;
- empty sum: $\sum_{i=a}^{b} s_i = 0$ if $b < a$;
- empty product: $\prod_{i=a}^{b} s_i = 1$ if $b < a$;
- Taylor series representation of a function, $f(x) \approx \sum_{n=0}^{\infty}(x-a)^n f^{(n)}(a)/n!$ (and the Maclaurin series, which is a Taylor Series expansion around $a = 0$);
- explicit definition of a function and implicit definition of a function;
- abbreviations for therefore $\therefore$ and because $\because$ and "for all" $\forall$ and "there exists" $\exists$ and "such that" or "subject to" s.t. ;
- sufficient condition, $\Longrightarrow$, necessary condition, $\Longleftarrow$, necessary and sufficient condition, $\Longleftrightarrow$, "iff" ("if and only if"; equivalence);[1]
- $\neg$ or $\sim$ to denote logical negation ("not") (though $\sim$ can also be used as a synonym for $\approx$, "approximately equal to")[2];
- converse;
- contrapositive;
- the terms "or" and "and" in mathematics ("or" is always understood to be the "inclusive or," meaning that 'A or B' is false if both A and B are false but it is true in all other cases, and in particular it is true if both A and B are true; 'A and B' is false unless both A and B are true; "exclusive or," or "xor," we will

[1] It is fine to write, for example, "$2x = 17 \Longrightarrow x = 17/2$." Never instead write the nonsensical "$2x = 17 = x = 17/2$." In other words, never confuse $\Longrightarrow$ and $=$.

[2] Do not use imprecise words such as "opposite" or "inverse" when you mean "negation."

not use in this course, but in case you are curious, it is false if both A and B are true (unlike 'A or B'), and it is false if both A and B are false; 'A xor B' is true if and only if exactly one of A or B is true);

- proof by contradiction (proving A by showing that "not A" implies a contradiction;
- proof by induction (prove a statement true for a small integer n_0; assume it true either for one larger integer n_1 ("weak induction") or, equivalently, for all integers in $[n_0, n_1]$ ("strong induction"); then prove the statement true for $n_1 + 1$);
- Axiom, Assumption, Theorem, Proposition, Lemma, Corollary;
- "Q.E.D." (Latin, "quod erat demonstrandum"), meaning "which had to be demonstrated;" also signified by $\square$ or by $\blacksquare$ or by $//$ or by $////$;
- the symbol $\approx$ (which is the most common symbol for "is approximately equal to" in the U.S.A.) and the symbol $\cong$ (which in the International Unicode standard means "is approximately equal to" but which in the U.S.A. is usually used in geometry to denote congruence (and in graph theory to denote isomorphic groups));
- homogeneous functions of degree k (namely, $f(\lambda\mathbf{x}) = \lambda^k f(\mathbf{x})$);
- homothetic functions (though we will review these).

Section 1:

- matrix, symmetric matrix;
- matrix transposition (denoted $\mathbf{A}^T$ or $\mathbf{A}'$);
- matrix determinant, $|\mathbf{A}|$ (vs. absolute value of a scalar) (note that $\begin{vmatrix} a & 2 \\ 5a & 6 \end{vmatrix} = a \begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix}$ and $\begin{vmatrix} a & 2a \\ 5a & 6a \end{vmatrix} = a^2 \begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix}$ whereas $\begin{bmatrix} a & a2 \\ 5a & 6a \end{bmatrix} = a \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix}$);
- C^n function;
- $f : \mathbf{R}^n \to \mathbf{R}^1$, domain, range, image of x under f, f maps its domain into its range, f is a mapping from its domain into its range (the ideas of a mapping being "one-to-one" or "onto" will be explained if those ideas are needed);
- notation such as f_3' and f_{42}'';
- gradient vector $\nabla f(\mathbf{x})$, Hessian matrix $\nabla^2 f(\mathbf{x})$;

- Jacobian matrix (see (12) below; it is square, but Jacobian matrices do not have to be square);
- linear combination of two vectors (e.g., $\alpha\mathbf{x} + \beta\mathbf{y}$);
- convex combination of two vectors (e.g., $\alpha\mathbf{x} + \beta\mathbf{y}$ with $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta = 1$);
- convex set;
- convex function, strictly convex function (namely $f''(x) > 0$ except possibly on a set of measure zero, where $f''(x) = 0$, e.g., $f(x) = x^4$), concave function, strictly concave function;
- quadratic form (for example, $f(x_1, x_2, x_3) = x_1x_2 - 3x_1x_3 + 4x_2x_2 = [x_1\ x_2\ x_3] \begin{bmatrix} 0 & 1/2 & -3/2 \\ 1/2 & 4 & 0 \\ -3/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$), positive definite, positive semidefinite, negative definite, negative semidefinite;
- contour line, upper level set (upper contour set), lower level set (lower contour set), quasiconcavity, strict quasiconcavity, quasiconvexity, strict quasiconvexity.

  The usual definition of quasiconcavity of f is that f's upper level sets are convex sets, but an equivalent definition is that $u(\alpha x + (1-\alpha)\,y) \geq \min\{u(x), u(y)\}\ \forall\ x,\ y$, and $\alpha \in [0, 1]$. This enables one to define the *strict* quasiconcavity of f as satisfying the conditions for quasiconcavity and also satisfying $u(\alpha x + (1-\alpha)\,y) > \min\{u(x), u(y)\}\ \forall\ x \neq y$ and $\alpha \in (0, 1)$. The level curve for a strictly quasiconcave function $f(\mathbf{x})$ is a strictly convex function of any one of the x_i, holding the other components of $\mathbf{x}$ fixed.

Section 2: admissible point, local minimum point, local maximum point, local minimum value, local maximum value, extreme points, extreme values; the * notation as traditionally denoting optima.

Section 3: binding and nonbinding inequality constraints, strict local minimum (namely x^* such that $f(x) > f(x^*)$ (strictly) for all x in a neighborhood of x^*), strict local maximum.

Section 4: global minimum (namely x^* such that $f(x) \geq f(x^*)$ for all x), global maximum, unique global minimum (namely x^* such that $f(x) > f(x^*)$ for all $x \neq x^*$), unique global maximum.

Section 5: endogenous variables, exogenous variables, dependent variables, independent variables, differential of a function of multiple variables, matrix inverse, Cramer's Rule.

Section 6: no additional new terms or notation.

Section 7: probability of an event, "$\{x : f(x) = 6\}$," "$\sum_i x_i$," "$\int f(x)\,dx$," "fair" random process.

1. Convexity, Quadratic Forms, and Minors

Let **A** denote a matrix. It does not have to be square. A "**minor** of **A** of order r" is obtained by deleting all but r rows and r columns of **A**, then taking the determinant of the resulting $r \times r$ matrix.

Now let **A** denote a square matrix. A "**principal minor** of **A** of order r" is obtained by deleting all but r rows and the *corresponding* r columns of **A**, then taking the determinant of the resulting $r \times r$ matrix. (For example, if you keep the first, third, and fourth rows, then you have to keep the first, third, and fourth columns.) A principal minor of **A** of order r is denoted by Δ_r of **A**.

Again let **A** denote a square matrix. A "**leading principal minor** of **A** of order r" is obtained by deleting all but the *first* r rows and the *first* r columns of **A**, then taking the determinant of the resulting $r \times r$ matrix. A leading principal minor of **A** of order r is denoted by D_r of **A**. A square matrix of dimension $n \times n$ has only 1 leading principal minor of order r for $r = 1, \ldots, n$.

Example. Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$. This matrix is not symmetric. Usually one is interested in the minors only of symmetric matrices, but there is nothing wrong with finding the minors of this non-symmetric matrix.

- The leading principal minor of order 1 of **A** is $D_1 = |1|$.

  There are four principal minors of order 1 of **A**; they are the Δ_1's: $|1| = D_1$, $|6|$, $|11|$, and $|16|$.

  There are sixteen minors of **A** of order 1.

- The leading principal minor of order 2 of **A** is $D_2 = \begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix}$.

  There are six principal minors of order 2 of **A**; they are the Δ_2's: $\begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix} = D_2$ (from rows and columns 1 and 2), $\begin{vmatrix} 1 & 3 \\ 9 & 11 \end{vmatrix}$ (from rows and columns 1 and 3), $\begin{vmatrix} 1 & 4 \\ 13 & 16 \end{vmatrix}$ (from rows and columns 1 and 4), $\begin{vmatrix} 6 & 7 \\ 10 & 11 \end{vmatrix}$ (from rows and columns 2 and 3), $\begin{vmatrix} 6 & 8 \\ 14 & 16 \end{vmatrix}$ (from rows and columns 2 and 4), and $\begin{vmatrix} 11 & 12 \\ 15 & 16 \end{vmatrix}$ (from rows and columns 3 and 4).

There are thirty-six minors of $\mathbf{A}$ of order 2.

- The leading principal minor of order 3 of $\mathbf{A}$ is $D_3 = \begin{vmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{vmatrix}$.

  There are four principal minors of order 3 of $\mathbf{A}$; they are the Δ_3's: $\begin{vmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{vmatrix} = D_3$ (from rows and columns 1, 2, and 3), $\begin{vmatrix} 1 & 2 & 4 \\ 5 & 6 & 8 \\ 13 & 14 & 16 \end{vmatrix}$ (from rows and columns 1, 2 and 4), $\begin{vmatrix} 1 & 3 & 4 \\ 9 & 11 & 12 \\ 13 & 15 & 16 \end{vmatrix}$ (from rows and columns 1, 3 and 4), and $\begin{vmatrix} 6 & 7 & 8 \\ 10 & 11 & 12 \\ 14 & 15 & 16 \end{vmatrix}$ (from rows and columns 2, 3 and 4).

  There are sixteen minors of $\mathbf{A}$ of order 3.

- The leading principal minor of order 4 of $\mathbf{A}$ is $D_4 = |\mathbf{A}|$.

  There is only one principal minor of order 4 of $\mathbf{A}$; it is Δ_4 and it is equal to $|\mathbf{A}|$.

  There is only one minor of order 4 of $\mathbf{A}$; it is $|\mathbf{A}|$.

[End of Example]

Let f be a C^2 function mapping $S \subset R^n$ into R^1. Denote the Hessian matrix of $f(\mathbf{x})$ by $\nabla^2 f(\mathbf{x})$; this matrix has dimension $n \times n$. Let "D_r of $\nabla^2 f(\mathbf{x})$" denote the rth-order leading principal minor of the Hessian of f. Let "Δ_r of $\nabla^2 f(\mathbf{x})$" denote all the rth-order principal minors of the Hessian of f.

Proposition 1. *One has*

$$
\begin{aligned}
& D_r \textit{ of } \nabla^2 f(\mathbf{x}) > 0 \textit{ for } r = 1, \ldots, n \textit{ and for all } \mathbf{x} \in S && (1)\\
& \iff \nabla^2 f(\mathbf{x}) \textit{ is positive definite for all } \mathbf{x} \in S && (2)\\
& \implies f(\mathbf{x}) \textit{ is strictly convex on } S\,. && (3)
\end{aligned}
$$

Also,

$$
\begin{aligned}
& \textit{All the } \Delta_r \textit{ of } \nabla^2 f(\mathbf{x}) \geq 0 \textit{ for } r = 1, \ldots, n \textit{ and for all } \mathbf{x} \in S && (4)\\
& \iff \nabla^2 f(\mathbf{x}) \textit{ is positive semidefinite for all } \mathbf{x} \in S && (5)\\
& \iff f(\mathbf{x}) \textit{ is convex on } S\,. && (6)
\end{aligned}
$$

If $\nabla^2 f(\mathbf{x})$ is replaced by an arbitrary symmetric matrix, it is still true that (1) $\iff$ (2) and (4) $\iff$ (5).

As a simple example that (3) implies neither (2) nor (1) (the implication only goes in the other direction), note that if $f(x) = x^4$ and if S is the entire real line, then since one possible value of x is zero (at which $\nabla^2 f(x) = 12x^2$ equals zero), (1)–(6) are, respectively, False, False, True, True, True, and True.

The typical procedure is to check (1) first. If (1) doesn't apply because one of the signs was strictly negative, then the contrapositive of "(4) iff (6)" tells you that the function is not convex. (This is because each $D_i \in \Delta_i$.) If (1) doesn't apply because at least one of the signs was zero but none were strictly negative, then one would have to check (4). The easiest part of (4) to check is the Δ_1's, which are the elements on the main diagonal of $\nabla^2 f(\mathbf{x})$. If any of them are strictly negative, then (4) fails, so (5) and (6) fail. (I may show you a direct proof in class that if a matrix is positive semidefinite, all its diagonal terms are greater than or equal to zero, and if a matrix is positive definite, all its diagonal terms are greater than zero.)

[Note that if f is convex then $-f$ is concave. This leads to:
Proposition 1′: Similarly,

$$\begin{aligned} &D_r \text{ of } \nabla^2 f(\mathbf{x}) \text{ alternate in sign beginning with } < 0 \text{ for } r = 1, \ldots, n \\ &\qquad \text{and } \forall\, \mathbf{x} \in S && (1') \\ &\iff \nabla^2 f(\mathbf{x}) \text{ is negative definite for all } \mathbf{x} \in S && (2') \\ &\implies f(\mathbf{x}) \text{ is strictly concave on } S\,. && (3') \end{aligned}$$

Also,

$$\begin{aligned} &\text{All the } \Delta_r \text{ of } \nabla^2 f(\mathbf{x}) \text{ alternate in sign beginning with } \leq 0 \text{ for } r = 1, \ldots, n \\ &\qquad \text{and } \forall\, \mathbf{x} \in S && (4') \\ &\iff \nabla^2 f(\mathbf{x}) \text{ is negative semidefinite for all } \mathbf{x} \in S && (5') \\ &\iff f(\mathbf{x}) \text{ is concave on } S\,. && (6') \end{aligned}$$

]

The following proposition is a test for "pseudoconvexity" and "pseudoconcavity," but for all practical purposes you should assume that pseudoconvexity is the same as quasiconvexity and pseudoconcavity is the same as quasiconcavity, so I will not even bother to define pseudoconvexity and pseudoconcavity.

Proposition 2. ***[Test of Pseudoconvexity.]*** *Let f be a C^2 function defined in an open, convex set S in R^n. Define the "bordered Hessian" determinants*

$\delta_r(\mathbf{x})$, $r = 1, \ldots, n$ by

$$\delta_r(\mathbf{x}) = \begin{vmatrix} 0 & f_1'(\mathbf{x}) & f_2'(\mathbf{x}) & \cdots & f_r'(\mathbf{x}) \\ f_1'(\mathbf{x}) & f_{11}''(\mathbf{x}) & f_{12}''(\mathbf{x}) & \cdots & f_{1r}''(\mathbf{x}) \\ f_2'(\mathbf{x}) & f_{21}''(\mathbf{x}) & f_{22}''(\mathbf{x}) & \cdots & f_{2r}''(\mathbf{x}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_r'(\mathbf{x}) & f_{r1}''(\mathbf{x}) & f_{r2}''(\mathbf{x}) & \cdots & f_{rr}''(\mathbf{x}) \end{vmatrix}.$$

A sufficient condition for f to be pseudoconvex is that $\delta_r(\mathbf{x}) < 0$ for $r = 2, \ldots, n$, and all $\mathbf{x} \in S$.

[Proposition 2′: Similarly, a sufficient condition for f to be pseudoconcave is that $\delta_r(\mathbf{x})$ alternate in sign beginning with > 0 for $r = 2, \ldots, n$, and all $\mathbf{x} \in S$.]

2. First-Order Conditions

Proposition 3. *Suppose that f, $h_1, \ldots, h_j$, and $g_1, \ldots, g_k$ are C^1 functions of n variables. Suppose that $\mathbf{x}^* \in R^n$ is a local minimum of $f(\mathbf{x})$ on the constraint set defined by the j equalities and k inequalities*

$$h_1(\mathbf{x}) = 0, \quad \ldots, \quad h_j(\mathbf{x}) = 0, \tag{7}$$

$$g_1(\mathbf{x}) \geq 0, \quad \ldots, \quad g_k(\mathbf{x}) \geq 0. \tag{8}$$

If there are no equality constraints then $j = 0$, and if there are no inequality constraints then $k = 0$. Form the Lagrangian

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{i=1}^{j} \lambda_i \, h_i(\mathbf{x}) - \sum_{i=1}^{k} \mu_i \, g_i(\mathbf{x}). \tag{9}$$

(If $j = 0$ or $k = 0$, recall the convention for empty sums. If both j and k are zero, that convention implies that $\mathscr{L} = f$.)

Then (under certain conditions I omit here) there exist multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that:

1. *$\partial \mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial \lambda_i = 0$ for all $i = 1, \ldots, j$. This is equivalent to: $h_i(\mathbf{x}^*) = 0$ for all $i = 1, \ldots, j$.*
2. *$\partial \mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial x_i = 0$ for all $i = 1, \ldots, n$.*
3. *$\mu_i^* \geq 0$, $g_i(\mathbf{x}^*) \geq 0$, and $\mu_i^* \, g_i(\mathbf{x}^*) = 0$ for all $i = 1, \ldots, k$.*

These three conditions are often called the Kuhn-Tucker conditions. The last condition is sometimes called the "complementary slackness condition."

[Proposition 3′: For a maximum, change (8) to:

$$g_1(\mathbf{x}) \le 0, \quad \dots, \quad g_k(\mathbf{x}) \le 0. \tag{8′}$$

Then the Lagrangian is formed in the same way. Condition 1 is unchanged. Condition 2 is unchanged. Condition 3 becomes: $\mu_i^* \ge 0$, $g_i(\mathbf{x}^*) \le 0$, and $\mu_i^* \, g_i(\mathbf{x}^*) = 0$ for all $i = 1, \dots, k$.]

In (7), if the first constraint of a problem is, for example, $x_1 + x_2 = 17$, one could either take $h_1(\mathbf{x}) = x_1 + x_2 - 17$ or take $h_1(\mathbf{x}) = 17 - x_1 - x_2$. Your choice will not affect the optimal value of $\mathbf{x}$. Your choice *will* affect the *sign* of $\boldsymbol{\lambda}^*$, but that sign is almost always unimportant. In the one case when it is important, equation (19) below, the discussion preceding equation (19) requires that the constraint be written in the particular order specified by equation (18).

3. Second-Order Conditions: Local

Let $\mathscr{L}$ be the Lagrangian of the optimization problem. In Section 2, I named the Lagrange multipliers "λ" if they were associated with one of the j equality constraints and "μ" if they were associated with one of the k inequality constraints. In this section: (a) ignore all the nonbinding inequality constraints at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$; and (b) rename the Lagrange multipliers of the binding inequality constraints $\lambda_{j+1}, \lambda_{j+2}, \dots, \lambda_m$, where:

> m is the number of equality constraints plus the number of binding inequality constraints.

(Do not confuse this use of m with Varian's textbook's use of m as standing for income.) It is allowed to have $m = 0$; if $m = 0$ then there are no Lagrange multipliers. Denote the m binding Lagrange multipliers collectively by $\boldsymbol{\lambda}$. Let there be n variables with respect to which the optimization is occurring; denote these variables collectively by $\mathbf{x}$.

A function's Hessian is not unique. For example, one Hessian of $f(x_1, x_2)$ is $\begin{bmatrix} f''_{11} & f''_{12} \\ f''_{21} & f''_{22} \end{bmatrix}$ and another is $\begin{bmatrix} f''_{22} & f''_{21} \\ f''_{12} & f''_{11} \end{bmatrix}$: the first one shows differentiation first with respect to x_1 and then with respect to x_2, while the second shows differentiation first with respect to x_2 and then with respect to x_1. Let $\nabla^2\mathscr{L}$ be the following particular Hessian of the Lagrangian: first differentiate

$\mathscr{L}$ with respect to all the Lagrange multipliers, then differentiate it with respect to the original variables $\mathbf{x}$.

$$\nabla^2\mathscr{L} = \begin{pmatrix} \mathscr{L}''_{\lambda_1\lambda_1} & \mathscr{L}''_{\lambda_1\lambda_2} & \cdots & \mathscr{L}''_{\lambda_1\lambda_m} & | & \mathscr{L}''_{\lambda_1 x_1} & \mathscr{L}''_{\lambda_1 x_2} & \cdots & \mathscr{L}''_{\lambda_1 x_n} \\ \mathscr{L}''_{\lambda_2\lambda_1} & \mathscr{L}''_{\lambda_2\lambda_2} & \cdots & \mathscr{L}''_{\lambda_2\lambda_m} & | & \mathscr{L}''_{\lambda_2 x_1} & \mathscr{L}''_{\lambda_2 x_2} & \cdots & \mathscr{L}''_{\lambda_2 x_n} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ \mathscr{L}''_{\lambda_m\lambda_1} & \mathscr{L}''_{\lambda_m\lambda_2} & \cdots & \mathscr{L}''_{\lambda_m\lambda_m} & | & \mathscr{L}''_{\lambda_m x_1} & \mathscr{L}''_{\lambda_m x_2} & \cdots & \mathscr{L}''_{\lambda_m x_n} \\ — & — & — & — & — & — & — & — & — \\ \mathscr{L}''_{x_1\lambda_1} & \mathscr{L}''_{x_1\lambda_2} & \cdots & \mathscr{L}''_{x_1\lambda_m} & | & \mathscr{L}''_{x_1 x_1} & \mathscr{L}''_{x_1 x_2} & \cdots & \mathscr{L}''_{x_1 x_n} \\ \mathscr{L}''_{x_2\lambda_1} & \mathscr{L}''_{x_2\lambda_2} & \cdots & \mathscr{L}''_{x_2\lambda_m} & | & \mathscr{L}''_{x_2 x_1} & \mathscr{L}''_{x_2 x_2} & \cdots & \mathscr{L}''_{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ \mathscr{L}''_{x_n\lambda_1} & \mathscr{L}''_{x_n\lambda_2} & \cdots & \mathscr{L}''_{x_n\lambda_m} & | & \mathscr{L}''_{x_n x_1} & \mathscr{L}''_{x_n x_2} & \cdots & \mathscr{L}''_{x_n x_n} \end{pmatrix}$$

If you do this right, $\nabla^2\mathscr{L}$ should have an $m \times m$ zero matrix in its upper left-hand corner:

$$\nabla^2\mathscr{L} = \begin{pmatrix} \mathscr{L}''_{\boldsymbol{\lambda\lambda}} & \mathscr{L}''_{\boldsymbol{\lambda x}} \\ \mathscr{L}''_{\boldsymbol{x\lambda}} & \mathscr{L}''_{\boldsymbol{xx}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathscr{L}''_{\boldsymbol{\lambda x}} \\ \mathscr{L}''^{\,T}_{\boldsymbol{\lambda x}} & \mathscr{L}''_{\boldsymbol{xx}} \end{pmatrix}$$

where $\mathscr{L}''_{\boldsymbol{\lambda x}}$ is an $m \times n$ matrix and where a 'T' superscript denotes the transpose.

One has the following result:

Proposition 4. *A sufficient condition for the point* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *identified in Proposition 3 to be a strict local minimum is that* $(-1)^m$ *has the same sign as all of the following when they are evaluated at* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$*:* D_{2m+1} *of* $\nabla^2\mathscr{L}$*,* D_{2m+2} *of* $\nabla^2\mathscr{L}, \ldots, D_{m+n}$ *of* $\nabla^2\mathscr{L}$*.*

If $m = 0$, this is equivalent to the condition that $\nabla^2\mathscr{L}$ (which in such a case equals $\nabla^2 f(\mathbf{x})$) be *positive definite*, which occurs only if $f(\mathbf{x})$ is strictly convex.

[Proposition 4$'$: Similarly, one will have a strict local maximum if, when they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, the following alternate in sign beginning with the sign of $(-1)^{m+1}$: D_{2m+1} of $\nabla^2\mathscr{L}$, D_{2m+2} of $\nabla^2\mathscr{L}, \ldots, D_{m+n}$ of $\nabla^2\mathscr{L}$.]

There is a second-order necessary condition for a minimum, also:

Proposition 5. *If* $m > 0$*, define "*$\widehat{\Delta}_i$ *of* $\nabla^2\mathscr{L}$*" to be the subset of "*Δ_i *of* $\nabla^2\mathscr{L}$*" formed by only considering those "*Δ_i *of* $\nabla^2\mathscr{L}$*" which retain (parts of) the first* m *rows and first* m *columns of* $\nabla^2\mathscr{L}$*. (If* $m = 0$*, there is no difference between the* Δ*'s and the* $\widehat{\Delta}$*'s.)*

Then a necessary condition for the point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *identified in Proposition 3 to be a local minimum is that "*$(-1)^m$ *or zero" have the same sign as all of the following when they are evaluated at* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$*:* $\widehat{\Delta}_{2m+1}$ *of* $\nabla^2 \mathscr{L}$*,* $\widehat{\Delta}_{2m+2}$ *of* $\nabla^2 \mathscr{L}, \ldots, \widehat{\Delta}_{m+n}$ *of* $\nabla^2 \mathscr{L}$*.*

The typical procedure is to check Proposition 4 first. If Proposition 4 doesn't apply because one of the signs was strictly the same as $(-1)^{m+1}$, then Proposition 5 tells you that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is not a local minimum point. (This is because each $D_i \in \widehat{\Delta}_i$.)

[Proposition 5′: The version of Proposition 5 for a local maximum requires that the following, if they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, alternate in sign beginning with the sign of "$(-1)^{m+1}$ or zero" (then having the sign of "$(-1)^{m+2}$ or zero" and so forth): $\widehat{\Delta}_{2m+1}$ of $\nabla^2 \mathscr{L}$, $\widehat{\Delta}_{2m+2}$ of $\nabla^2 \mathscr{L}, \ldots, \widehat{\Delta}_{m+n}$ of $\nabla^2 \mathscr{L}$.]

4. Second-Order Conditions: Global

Let $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ be a point identified in Proposition 3. Let j be the number of equality constraints and k be the number of inequality constraints.

1. If $j = k = 0$ (an **unconstrained** problem) and $f(\mathbf{x})$ is convex for all x, then $\mathbf{x}^*$ is a **global minimum** point of f in S. (The converse also holds.) Furthermore, if $j = k = 0$ and $f(\mathbf{x})$ is *strictly* convex for all x, then $\mathbf{x}^*$ is the **unique** global minimum point of f in S. (The converse also holds.)

2. If $k = 0$ (only **equality** constraints) and if $\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda})$ is "convex in $\mathbf{x}$" (that is, $\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda})$ is convex when considering all the components of $\boldsymbol{\lambda}$ to be constants instead of variables), then $\mathbf{x}^*$ is a **global constrained minimum** point of f. Furthermore, if $k = 0$ and $\mathscr{L}(\mathbf{x})$ is *strictly* convex in $\mathbf{x}$, then $\mathbf{x}^*$ is the **unique** global constrained minimum point of f.

[Aside: Similarly, using the problem defined in Proposition 3′ and (8′):

1′. If $j = k = 0$ (an **unconstrained** problem) and $f(\mathbf{x})$ is concave, then $\mathbf{x}^*$ is a **global maximum** point of f in S. (The converse also holds.) Furthermore, if $j = k = 0$ and $f(\mathbf{x})$ is *strictly* concave, then $\mathbf{x}^*$ is the **unique** global maximum point of f in S. (The converse also holds.)

2′. If $k = 0$ (only **equality** constraints) and $\mathscr{L}(\mathbf{x})$ is concave in $\mathbf{x}$, then $\mathbf{x}^*$ is a **global constrained maximum** point of f. Furthermore, if $k = 0$ and $\mathscr{L}(\mathbf{x})$ is *strictly* concave in $\mathbf{x}$, then $\mathbf{x}^*$ is the **unique** global constrained maximum point of f.

]

5. Comparative Statics

Let $\mathbf{x} \in \mathbf{R}^n$ denote the endogenous (or "dependent") variables in a model and let $\mathbf{y} \in \mathbf{R}^m$ denote the exogenous (or "independent") variables in that model. Suppose the model is described by a general system of equations of the form

$$\begin{aligned} f_1(\mathbf{x},\mathbf{y}) &= 0 \\ f_2(\mathbf{x},\mathbf{y}) &= 0 \\ &\vdots \\ f_n(\mathbf{x},\mathbf{y}) &= 0. \end{aligned} \tag{10}$$

This is called the "structural form" because it defines $\mathbf{x}$ as an *implicit* function of $\mathbf{y}$; if one could solve the system for $\mathbf{x}$ as an *explicit* function of $\mathbf{y}$, one would obtain the "reduced form" of the system. Often it is impossible to solve for the reduced form.

Taking the differential of both sides of each equation results in[3]

$$\begin{aligned} \frac{\partial f_1}{\partial x_1}dx_1 + \cdots + \frac{\partial f_1}{\partial x_n}dx_n + \frac{\partial f_1}{\partial y_1}dy_1 + \cdots + \frac{\partial f_1}{\partial y_m}dy_m &= 0 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \quad & \\ \frac{\partial f_n}{\partial x_1}dx_1 + \cdots + \frac{\partial f_n}{\partial x_n}dx_n + \frac{\partial f_n}{\partial y_1}dy_1 + \cdots + \frac{\partial f_n}{\partial y_m}dy_m &= 0. \end{aligned} \tag{11}$$

Moving the last m terms in each equation to the right (in order to isolate the differentials of the endogenous variables, so those differentials can be solved for), and rewriting in matrix form, results in

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix} = -\begin{pmatrix} \frac{\partial f_1}{\partial y_1} \\ \vdots \\ \frac{\partial f_n}{\partial y_1} \end{pmatrix} dy_1 - \cdots - \begin{pmatrix} \frac{\partial f_1}{\partial y_m} \\ \vdots \\ \frac{\partial f_n}{\partial y_m} \end{pmatrix} dy_m\,. \tag{12}$$

Let $\mathbf{J}$ denote the matrix on the left-hand side of (12). (This matrix is a "Jacobian matrix.") If $\mathbf{J}$ is invertible then we can solve for dx_1, dx_2, ...,

[3]Do not omit the " $= 0$" parts of (11). If you omit them, you have only taken the differential of one side of (10) instead of both sides of (10), and you would not be able to make any further progress.

dx_n as a function of $dy_1, dy_2, \ldots, dy_m$ as follows:

$$\begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix} = -\mathbf{J}^{-1} \begin{pmatrix} \dfrac{\partial f_1}{\partial y_1} \\ \vdots \\ \dfrac{\partial f_n}{\partial y_1} \end{pmatrix} dy_1 - \cdots - \mathbf{J}^{-1} \begin{pmatrix} \dfrac{\partial f_1}{\partial y_m} \\ \vdots \\ \dfrac{\partial f_n}{\partial y_m} \end{pmatrix} dy_m \,. \tag{13}$$

Alternatively and more commonly, (12) is solved using Cramer's Rule, especially in the many problems in which most of the dy's are zero.

Consider the common problem of determining the sign of $\partial x_i / \partial y_j$ from (12). The easiest way to do this, if (10) are the first-order conditions of an optimization problem (which in microeconomics is usually the case), is usually to solve (12) using Cramer's Rule. Then $\partial x_i / \partial y_j$ would have the form

$$\frac{\text{numerator}}{|\mathbf{J}|} \,. \tag{14}$$

In such cases, the second-order conditions of the optimization problem usually determine the sign of $|\mathbf{J}|$; then all that remains in order to determine the sign of dx_i / dy_j is to find the sign of the numerator of (14).

6. The Value Function and the Envelope Theorem

Consider the problem of maximizing a function f over endogenous variables $\mathbf{x}$ given exogenous variables $\mathbf{c}$ and constraints $h_1(\mathbf{x}, \mathbf{c}) = 0$, $h_2(\mathbf{x}, \mathbf{c}) = 0$, $\ldots$, $h_j(\mathbf{x}, \mathbf{c}) = 0$. The "(optimized) value function" for this problem is defined as

$$\begin{aligned} f^*(\mathbf{c}) = \max_{\boldsymbol{x}} f(\mathbf{x}, \mathbf{c}) \quad & \text{such that} \\ h_1(\mathbf{x}, \mathbf{c}) = 0, \quad h_2(\mathbf{x}, \mathbf{c}) = 0, \quad & \ldots, \quad h_j(\mathbf{x}, \mathbf{c}) = 0. \end{aligned} \tag{15}$$

Equivalently, if $\mathbf{x}^*$ is the solution to the maximization problem in (15), then

$$f^*(\mathbf{c}) = f(\mathbf{x}^*(\mathbf{c}), \mathbf{c}) \,. \tag{16}$$

Let $\mathscr{L}$ be the Lagrangian function (9) for the maximization problem in (15). The "Envelope Theorem" states that

$$\frac{\partial f^*}{\partial c_i} = \frac{\partial \mathscr{L}^*}{\partial c_i} \tag{17}$$

where $\mathscr{L}^*$ is $\mathscr{L}$ evaluated at $(\mathbf{x}^*, \mathbf{c})$.

Consider the special case of (15) in which $f(\mathbf{x}, \mathbf{c})$ does not depend on $\mathbf{c}$ and in which the constraints take the form

$$h_1(\mathbf{x}) - c_1 = 0, \quad \ldots, \quad h_j(\mathbf{x}) - c_j = 0\,. \tag{18}$$

Then (17) implies that

$$\frac{\partial f^*}{\partial c_i} = \frac{\partial \mathscr{L}^*}{\partial c_i} = \lambda_i^*\,. \tag{19}$$

This is often used to give an interpretation of λ_i^* as a "shadow price" of c_i.

7. Probability Theory

All the probability theory that is required for this course is an understanding of how to compute the expected value of a discrete or continuous random variable. A superficial understanding will suffice, but some students might be interested in a more careful treatment, which I give below. However, I still will not be giving a fully satisfactory treatment, because that requires measure theory, Borel sets, and other advanced mathematics; such a treatment is given for example in Chapter 1 of Malliaris and Brock's 1982 textbook "Stochastic Methods in Economics and Finance."

Let the set of possible outcomes of an uncertain event be called the "sample space" and be denoted by Ω. We will first suppose that the number of elements in Ω is finite or countably infinite.

With each element $\omega \in \Omega$ associate a real number $X(\omega)$. For example, if Ω is a deck of playing cards and each ω is one card, then $X(\omega)$ might be 1 when ω is the 2 of Hearts, 10 when ω is the Jack of Hearts, 14 when ω is the 2 of Spades, and so forth. The function $X : \Omega \to \mathbf{R}$ is called a "discrete random variable."

Let $\Pr(\omega)$ denote the probability that ω occurs. Let the function $f(x) : \mathbf{R} \to [0, 1]$ be defined by

$$f(x) = \Pr\{\omega : X(\omega) = x\}\,.$$

The function f is called the "probability distribution" of the discrete random variable X. One has

$$\sum_{x \in \mathbf{R}} f(x) = 1\,.$$

The "expected value" of the random variable (also called the "mean" of the random variable or the "average" of the random variable) is defined to be

$$E(X) = \sum_{x \in \mathbf{R}} x f(x)\,.$$

For example, consider the outcome of a roll of a die. The set of outcomes, in no particular order, is $\Omega = \{3, 1, 5, 4, 6, 2\}$. Let the "first" outcome be $\omega_1 = 3$, the "second" outcome be $\omega_2 = 1$, and so forth, so the sixth outcome is $\omega_6 = 2$. Define the random variable $X(\omega)$ in the following way: $X(\omega_1) = 3^2 = x_1$, $X(\omega_2) = 1^2 = x_2$, $\ldots$, $X(\omega_6) = 2^2 = x_6$. If in addition the die is fair (so all the outcomes occur with probability $1/6$), then the expected value of X is

$$\begin{aligned}\sum_{i=1}^{6} x_i\, f(x_i) &= 3^2 \cdot \frac{1}{6} + 1^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} \\ &= \frac{1}{6} \cdot (9 + 1 + 25 + 16 + 36 + 4) = 91/6 = 15\tfrac{1}{6}\,.\end{aligned}$$

For another example, again consider the outcome of a roll of a die. This time write the set of outcomes as $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let the "first" outcome be $\omega_1 = 1$, the "second" outcome be $\omega_2 = 2$, and so forth, so the sixth outcome is $\omega_6 = 6$. Define the random variable $Y(\omega)$ in the following way: $Y(\omega_1) = 1 = y_1$, $Y(\omega_2) = 2 = y_2$, $\ldots$, $Y(\omega_6) = 6 = y_6$. If in addition the die is fair (so all the outcomes occur with probability $1/6$), then the expected value of Y is

$$\begin{aligned}\sum_{i=1}^{6} y_i\, f(y_i) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} \\ &= \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = 21/6 = 3.5\,.\end{aligned}$$

This completes our treatment of the case when the number of elements in Ω is finite or countably infinite. Now suppose instead that the number of elements in Ω is uncountably infinite.

Furthermore, suppose that to each element $\omega \in \Omega$ we can associate a real number $X(\omega)$. For example, if ω is the color of paint in a paint can which we find together with many other paint cans in an abandoned building, then Ω is the set of all possible colors in the abandoned cans, and if red is one's favorite color, then $X(\omega)$ might be the grams of red pigment contained in the first abandoned paint can. The function $X : \Omega \to \mathbf{R}$ is called a "continuous random variable."

Let the function $F(x) : \mathbf{R} \to [0, 1]$ be defined by

$$F(x) = \Pr\{\omega : X(\omega) \le x\}\,.$$

The function F is called the "cumulative probability density function," or CDF, of the continuous random variable X. (In the example, CDF(x) is the probability that the paint can will have less than or equal to x grams of red pigment.) One has $F(\infty) = 1$.

The function

$$f(x) = \frac{dF(x)}{dx}$$

is called the "probability density function," or PDF, of the continuous random variable X. One has

$$\int_{-\infty}^{\infty} f(x)\, dx = 1\,.$$

The probability that the value of X is between a and b (where $a \neq b$) is $\int_a^b f(x)\, dx$. The probability that the value of X is exactly equal to a is not given by $\int_a^a f(x)\, dx = 0$, because then X could never take on any value. Instead, frequency with which the value of X is exactly equal to any particular value "a" goes to zero in the limit as the number of draws from the distribution goes to infinity.

The "expected value" of the random variable (also called the "mean" of the random variable or the "average" of the random variable) is defined to be

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx\,.$$

For example, if $\Omega = [0, 12]$ for the outcome of the spin of a fair arrow centered on the face of a clock, if ω is defined be the number that the arrow points to on the clock face, and if $X(\omega)$ is defined to equal ω (so $X(3) = 3$), then the CDF of the arrow is 0.25 at $x = 3$, 0.75 at $x = 9$, and in general is equal to $x/12$. The PDF in this example is

$$f(x) = 1/12\,,$$

and the expected value is

$$\int_0^{12} x \cdot \frac{1}{12}\, dx = \frac{1}{12} \cdot \frac{1}{2} x^2 \Big|_0^{12} = 6\,.$$