London School of Economics
Department of Economics

Dr Francesco Nava
Office: 32L.3.20

EC400 – Math for Microeconomics
Syllabus 2019/20

The course is based on three lectures (2.5 hours each) and on three classes (2.5 hours each). Lectures revise mathematical notions essential for Microeconomics. Classes show how to apply these notions to solve relevant problems.

Course Outline

Lecture 1: Tools for Optimization

- Quadratic Forms
- Determinant
- Taylor's Expansion
- Concavity and Convexity
- Quasi-Concavity and Quasi-Convexity

Lecture 2: Optimization

- Unconstrained Optimization
- Constrained Optimization: Equality Constraints
- Lagrange Theorem
- Constrained Optimization: Inequality Constraints
- Kuhn-Tucker Theorem
- Arrow-Enthoven Theorem

Lecture 3: Comparative Statics and Fixed Points

- Envelope Theorem
- Implicit Function Theorem
- Correspondences
- Fixed Point Theorems
- Theorem of the Maximum

Suggested Textbooks

- Simon & Blume, *Mathematics for Economists*, Norton, 1994.
- Takayama, *Mathematical Economics*, Cambridge University Press, 1985.
- Chiang & Wainwright, *Fundamental Methods of Mathematical Economics*, McGrawHill, 2005.
- Dixit, *Optimization in Economic Theory*, Oxford University Press, 1990.
- Ok, *Real Analysis with Economic Applications*, Princeton University Press, 2007.
- Chiang, *Elements of Dynamic Optimization*, McGrawHill, 1992.
- Kamien & Schwartz, *Dynamic Optimization*, Elsevier Science, 1991.

Excercises

- Three problem sets to test background and preparation.
- Access to past exam paper to prepare for the exam.

London School of Economics
Department of Economics

Dr Francesco Nava
Office: 32L.3.20

EC400 – Math for Microeconomics
Lecture Notes 2019/20

Course Outline

Lecture 1: Tools for Optimization

- Quadratic Forms
- Determinant
- Taylor's Expansion
- Concavity and Convexity
- Quasi-Concavity and Quasi-Convexity

Lecture 2: Optimization

- Unconstrained Optimization
- Constrained Optimization: Equality Constraints
- Lagrange Theorem
- Constrained Optimization: Inequality Constraints
- Kuhn-Tucker Theorem
- Arrow-Enthoven Theorem

Lecture 3: Comparative Statics and Fixed Points

- Envelope Theorem
- Implicit Function Theorem
- Correspondences
- Fixed Point Theorems
- Theorem of the Maximum

Lecture 1: Tools for Optimization
Quadratic Forms and Their Determinant

Defining Quadratic Forms

Quadratic forms are useful and important because: (i) the simplest functions after linear ones; (ii) conditions for optimization techniques are stated in terms of quadratic forms; (iii) relevant economic optimization problems have a quadratic objective function, such as risk minimization problems in finance (where riskiness is measured by the quadratic variance of the returns from investments).

Definition: A *quadratic form* on R^n is a real valued function:

$$Q(x_1, x_2, ..., x_n) = \sum_{i \leq j} a_{ij} x_i x_j$$

Among all functions of one variable, the simplest functions with a unique global extremum are the pure quadratic forms: x^2 and $-x^2$. The level curve of a general quadratic form in $\mathbb{R}^2$ is

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 = b$$

such curve can take the form of an ellipse, a hyperbola, a pair of lines, or the empty set. A a quadratic form in $\mathbb{R}^2$ $a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$ can always be written as

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^T A \mathbf{x}.$$

Every quadratic form can be represented as

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

where A is a (unique) **symmetric** matrix,

$$A = \begin{pmatrix} a_{11} & a_{12}/2 & ... & a_{1n}/2 \\ a_{21}/2 & a_{22} & ... & a_{2n}/2 \\ ... & ... & ... & ... \\ a_{n1}/2 & a_{n2}/2 & ... & a_{nn} \end{pmatrix}.$$

Conversely if A is a symmetric matrix, then the real valued function $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, is a quadratic form.

Remark: Any quadratic form always takes the value 0 when $\mathbf{x} = \mathbf{0}$.

We want to know whether $\mathbf{x} = \mathbf{0}$ is a maximum, a minimum, or neither.

For example for the quadratic form ax^2:

1. if $a > 0$, ax^2 is non negative and equal to 0 only when $x = 0$ which implies that the quadratic form is positive definite and that $x = 0$ is a global minimizer;

2. if $a < 0$, then the function is negative definite and $x = 0$ is a global maximizer.

In two dimensions there are additional intermediate cases, the quadratic form is:

1. **positive definite** when it is always non-negative, and positive for all $\mathbf{x} \neq 0$, eg:
$$x_1^2 + x_2^2;$$

2. **negative definite** when it is always non-positive, and negative for all $\mathbf{x} \neq 0$, eg:
$$-(x_1 + x_2);$$

3. **indefinite** when it can take both positive and negative values, eg:
$$x_1^2 - x_2^2;$$

4. **positive semidefinite** when it is non-negative, but equal to zero for some $\mathbf{x} \neq 0$, eg:
$$(x_1 + x_2)^2;$$

5. **negative semidefinite** when it is non-positive, but equal to zero for some $\mathbf{x} \neq 0$, eg:
$$-(x_1 + x_2)^2.$$

We apply the same terminology for the symmetric matrix A and the implied quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

Definition: Let A be an $(n \times n)$ symmetric matrix. If so, A is:
(a) *positive definite* if $x^T Ax > 0$ for all $x \neq 0$ in R^n;
(b) *positive semidefinite* if $x^T Ax \geq 0$ for all $x \neq 0$ in R^n;
(c) *negative definite* if $x^T Ax < 0$ for all $x \neq 0$ in R^n;
(d) *negative semidefinite* if $x^T Ax \leq 0$ for all $x \neq 0$ in R^n;
(e) *indefinite* $x^T Ax > 0$ for some $x \neq 0$ in R^n and $x^T Ax < 0$ for some $x \neq 0$ in R^n.

Application: A function $y = f(x)$ of one variable is concave if its second derivative $f''(x) \leq 0$ on some interval. The generalization of this concept to higher dimensions states that a function is concave on some region if its second derivative matrix is negative semidefinite for all $\mathbf{x}$ on that region.

The determinant of a matrix is a unique scalar associated with the matrix. Rather than showing that the many usual definitions of the determinant are all the same by comparing them to each other, I'm going to state four general properties of the determinant that are enough to specify uniquely what number you should get when given a matrix. Then, you may check that all of the definitions for determinant that you've encountered satisfy those properties. What we're really looking for is a function that takes n vectors (the n columns of the matrix) and returns a number. Consider a matrix A, and denote its i^{th} column by A_i so that

$$A = (A_1, A_2, ..., A_n).$$

Assume we're working with real numbers. Standard operation change the value of the determinant, as follows:

1. switching two columns changes the sign,

$$\det A = -\det(A_2, A_1, A_3, ..., A_n);$$

2. multiplying one column by a constant multiplies the determinant by that constant,

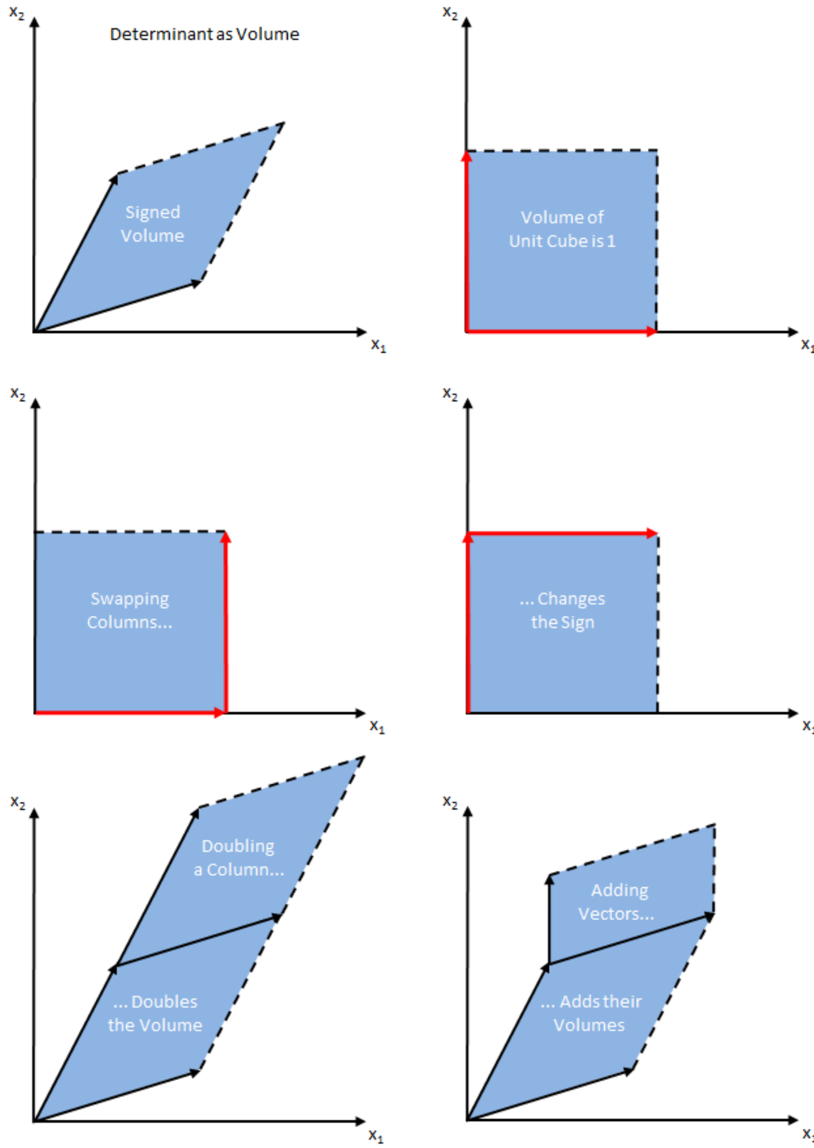
$$a \det A = \det(aA_1, A_2, ..., A_n);$$

3. the determinant is linear in each column,

$$\det(A_1 + B_1, A_2, ..., A_n) = \det(A_1, A_2, ..., A_n) + \det(B_1, A_2, ..., A_n).$$

These three properties, together with the fact that the determinant of the identity matrix is one, are enough to define a unique function that takes n vectors (each of length n) as inputs and returns a real number which amounts to the determinant of the matrix given by those vectors.

This observation helps with some of the interpretations of the determinant. In particular, there's a nice geometric way to think of a determinant. Consider the unit cube in n dimensional space (that is the set of vectors of length n with coordinates 0 or 1 in each spot). The determinant of the linear transformation (i.e. of the matrix) T is the signed volume of the region you get by applying T to the unit cube. Consider how that follows from the abstract definition. If you apply the identity to the unit cube, you get back the unit cube, and the volume of the unit cube is 1. If you stretch the cube by a constant factor in one direction only, the new volume is that constant. And if you stack two blocks together aligned on the same direction, their combined volume is the sum of their volumes: this all shows that the signed volume we have is linear in each coordinate when considered as a function of the input vectors. Finally, when you switch two of the vectors that define the unit cube, you flip the orientation.



With this geometric definition of determinant, if you are familiar with multivariate calculus, you could think about why determinants (the Jacobian) appears when we change coordinates doing integration. Hint: a derivative is a linear approximation of the associated function, and consider a "differential volume element" in your starting coordinate system. It's not too much work to check that the area of the parallelogram formed by vectors (a, b) and (c, d) is $\det((a, b), (c, d))$.

Computing the determinant ($\det A$ or equivalently $|A|$) of a matrix A, is straightforward in two dimensions,

$$\det A = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

and in three dimensions,

$$\det A = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} =$$
$$= a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

With more than three dimensions standard recursive definitions can be applied for its calculation. To do so, denote by:

- a_{ij} the entry in row i and column j of a matrix A;
- M_{ij} the submatrix obtained by removing row i and column j form a matrix A.

If so, *Laplace's formula* states that:

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij}.$$

Definition: Let A be an $(n \times n)$ matrix. Any $(k \times k)$ submatrix of A formed by deleting $(n-k)$ columns, say columns $(i_1, i_2, ..., i_{n-k})$ and the corresponding $(n-k)$ rows from A, $(i_1, i_2, ..., i_{n-k})$, is called a k^{th} order *principal submatrix* of A. The determinant of a k^{th} order principal submatrix is called a k^{th} order *principal minor* of A.

Definition: Let A be an $(n \times n)$ matrix. The k^{th} order principal submatrix of A obtained by deleting the *last* $(n-k)$ rows and columns from A is called the k^{th} order *leading principal submatrix* of A, denoted L_k. Its determinant is called the k^{th} order *leading principal minor* of A.

Example: A general (3×3) matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

possesses one third order principal minor, namely $\det A$, three second order principal minors,

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \det \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix},$$

and three first order principal minors a_{11}, a_{22}, a_{33}. The leading principal minors are

$$L_3 = \det A,\ L_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},\ L_1 = a_{11}.$$

Let A be an $(n \times n)$ symmetric matrix, then A is:

- positive definite if and only if **every leading principal minor** is strictly positive;

$$|L_1| > 0, \ |L_2| > 0, \ |L_3| > 0...$$

- negative definite if and only if every leading principal minor alternates in sign

$$|L_1| < 0, \ |L_2| > 0, \ |L_3| < 0...$$

  with the k^{th} order leading principal minor having the same sign as $(-1)^k$.

- positive semidefinite if and only if **every principal minor** of A is non-negative;

- negative semidefinite if and only if every principal minor of odd order is non-positive and every principal minor of even order is non-negative.

Example: Diagonal matrices correspond to the simplest quadratic forms

$$A = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix}.$$

Any diagonal quadratic form can be written as

$$\mathbf{x}^T A \mathbf{x} = a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2.$$

Obviously such a quadratic form will be: positive (negative) definite if and only if every a_i is positive (negative); positive (negative) semidefinite if and only if every a_i is non-negative (non-positive); indefinite if there are two entries a_i with opposite signs.

Example: Consider a (2×2) matrix A and its corresponding quadratic form:

$$Q(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a x_1^2 + 2b x_1 x_2 + c x_2^2$$

If $a = 0$, then Q cannot be definite since $Q(1, 0) = 0$. If $a \neq 0$ instead, add and subtract $b^2 x_2^2 / a$ to get

$$\begin{aligned} Q(x_1, x_2) &= a x_1^2 + 2b x_1 x_2 + c x_2^2 + \frac{b^2}{a} x_2^2 - \frac{b^2}{a} x_2^2 \\ &= a(x_1^2 + \frac{2b x_1 x_2}{a} + \frac{b^2}{a^2} x_2^2) - \frac{b^2}{a} x_2^2 + c x_2^2 \\ &= a(x_1 + \frac{b}{a} x_2)^2 + \frac{(ac - b^2)}{a} x_2^2. \end{aligned}$$

If both coefficients, a and $(ac-b^2)/a$, are positive, then Q will never be negative. It will equal 0 only when $x_1 + \frac{b}{a}x_2 = 0$ and $x_2 = 0$ in other words, when $x_1 = 0$ and $x_2 = 0$. That is, if

$$|a| > 0 \text{ and } \det A = (ac - b^2) > 0$$

then Q is positive definite. Conversely, if Q is positive definite then both a and $\det A = ac - b^2$ are positive.

Similarly, Q is negative definite if and only if both coefficient are negative, which occurs if and only if $a < 0$ and $ac - b^2 > 0$, that is, when the leading principal minors alternative in sign.

Finally, if $a < 0$ and $ac - b^2 < 0$, then the two coefficients will have opposite signs and Q will be indefinite.

Examples Testing Definiteness:

- Consider $A = \begin{pmatrix} 2 & 3 \\ 3 & 7 \end{pmatrix}$. Since $|A_1| = 2$ and $|A_2| = 5$, A is positive definite.
- Consider $B = \begin{pmatrix} 2 & 4 \\ 4 & 7 \end{pmatrix}$. Since $|B_1| = 2$ and $|B_2| = -2$, B is indefinite.
- Consider $Q(\mathbf{x}) = bx_1^2 + a(x_2^2 + x_3^2) + 2bx_2x_3$. Its principal submatrices are:

$$3^{\text{rd}}: \begin{pmatrix} b & 0 & 0 \\ 0 & a & b \\ 0 & b & a \end{pmatrix}; \ 2^{\text{nd}}: \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix}, \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix}, \begin{pmatrix} a & b \\ b & a \end{pmatrix}; \ 1^{\text{st}}: \ (b), \ (a), (a).$$

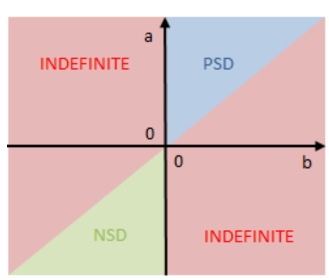
  $Q(\mathbf{x})$ is positive semidefinite if all principal minors are non-negative which requires:

$$a \geq 0, \ b \geq 0, \ ab \geq 0, \ a^2 - b^2 \geq 0, \ a^2b - b^3 \geq 0.$$

  For these to hold at once it must be that $a \geq 0$, $b \geq 0$, $a \geq b$. $Q(\mathbf{x})$ is negative semidefinite is all principal minors suitably alternate in sign which requires:

$$a \leq 0, \ b \leq 0, \ ab \geq 0, \ a^2 - b^2 \geq 0, \ a^2b - b^3 \leq 0.$$

  For these to hold at once it must be that $a \leq 0$, $b \leq 0$, $b \geq a$. Thus, the quadratic form is indefinite if and only if either $ab < b^2$ or $ab < 0$. The plot below depicts the answer



Lecture 1: Tools for Optimization
Taylor Theorem

Taylor's Expansion

The second tool needed for maximization is Taylor's expansion (or approximation). The first order Taylor's approximation of a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ satisfies

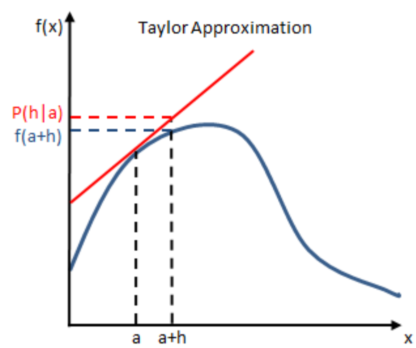
$$P(h|a) = f(a) + f'(a)h.$$

The polynomial approximates f around a, since $f(a+h)$ can be written as

$$f(a+h) = f(a) + f'(a)h + R(h|a)$$

where $R(h|a)$ is the difference between the two sides of the approximation.

The definition of the derivative $f'(a)$ implies that $\frac{R(h|a)}{h} \to 0$ as $h \to 0$.



Geometrically, this formalizes the approximation of f by its tangent line at $(a, f(a))$. Analytically, it describes the best approximation of f by a polynomial of degree 1.

Definition: The k^{th} *order Taylor expansion* of f at $x = a$ is

$$P_k(h|a) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \ldots + \frac{f^{[k]}(a)}{k!}h^k.$$

Moreover the corresponding Taylor residual $R_k(h|a)$ satisfies

$$R_k(h|a) = f(a+h) - P_k(h|a) \qquad \text{and} \qquad \lim_{h \to 0} \frac{R_k(h|a)}{h^k} = 0.$$

Example: Compute the first and second order Taylor polynomial of the exponential function $f(a) = e^a$ at $a = 0$. Since all the derivatives of f at $a = 0$ are equal 1:

$$P_1(h|0) = 1 + h$$
$$P_2(h|0) = 1 + h + \frac{h^2}{2}$$

For $h = 0.2$, then $P_1(0.2|0) = 1.2$ and $P_2(0.2|0) = 1.22$ compared with the actual value of $e^{0.2} = 1.2214$.

Multivariate Taylor's Expansion

For functions of several variables the first order Taylor expansion satisfies

$$P_1(\mathbf{h}|\mathbf{a}) = F(\mathbf{a}) + \frac{\partial F}{\partial x_1}(\mathbf{a})h_1 + ... + \frac{\partial F}{\partial x_n}(\mathbf{a})h_n = F(\mathbf{a}) + DF(\mathbf{a})\mathbf{h},$$

where $DF(\mathbf{a})$ denotes the *Jacobian*,

$$DF(\mathbf{a}) = \left(\frac{\partial F}{\partial x_1}(\mathbf{a}), ..., \frac{\partial F}{\partial x_n}(\mathbf{a})\right).$$

As before, the approximation holds since

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{R_1(\mathbf{h}|\mathbf{a})}{||\mathbf{h}||} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a}) - DF(\mathbf{a})\mathbf{h}}{||\mathbf{h}||} = 0.$$

The second order Taylor expansion satisfies

$$P_2(\mathbf{h}|\mathbf{a}) = F(\mathbf{a}) + DF(\mathbf{a})\mathbf{h} + \frac{1}{2!}\mathbf{h}^T D^2F(\mathbf{a})\mathbf{h},$$

where $D^2F(\mathbf{a})$ denotes the *Hessian*,

$$D^2F(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1 \partial x_1}(\mathbf{a}) & ... & \frac{\partial^2 F}{\partial x_n \partial x_1}(\mathbf{a}) \\ ... & ... & ... \\ \frac{\partial^2 F}{\partial x_1 \partial x_n}(\mathbf{a}) & ... & \frac{\partial^2 F}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix}.$$

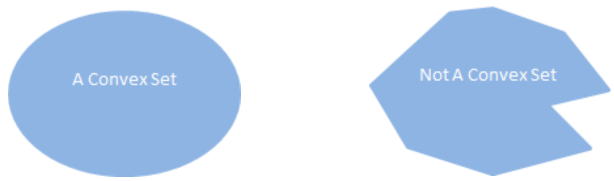
The extension for order k then follows accordingly.

Lecture 1: Tools for Optimization
Concavity and Quasi-Concavity

Concavity and Convexity

Definition: A set U is *convex* if for all $\mathbf{x}, \mathbf{y} \in U$ and for all $t \in [0,1]$

$$t\mathbf{x} + (1-t)\mathbf{y} \in U.$$

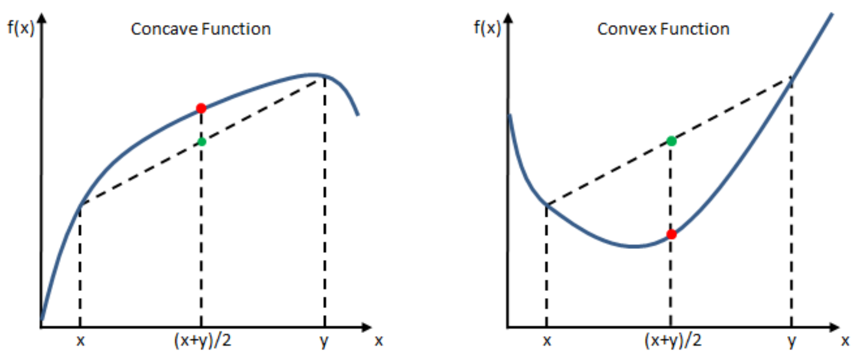


Definition: A real valued function f defined on a convex subset U of $\mathbb{R}^n$ is:
(1) *concave*, if for all $\mathbf{x}, \mathbf{y} \in U$ and for all $t \in [0,1]$

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \geq tf(\mathbf{x}) + (1-t)f(\mathbf{y});$$

(2) *convex*, if for all $\mathbf{x}, \mathbf{y} \in U$ and for all $t \in [0,1]$

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$



Observations:

(1) A function f is concave if and only if $-f$ is convex.

(2) Linear functions are both convex and concave.

(3) The definitions of concavity and convexity require f to have a convex set as a domain.

Claim: Let f be a differentiable function on a convex subset U of $\mathbb{R}^n$. If so, f is concave on U if and only if for all $\mathbf{x}, \mathbf{y} \in U$

$$f(\mathbf{y}) - f(\mathbf{x}) \leq Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) = \textstyle\sum_{i=1}^{n} \dfrac{\partial f(\mathbf{x})}{\partial x_i}(y_i - x_i).$$

Proof: The function $f : \mathbb{R} \to \mathbb{R}$ is concave if and only if

$$tf(y) + (1-t)f(x) \leq f(ty + (1-t)x) \quad \Leftrightarrow \quad t(f(y) - f(x)) + f(x) \leq f(x + t(y-x)) \quad \Leftrightarrow$$
$$f(y) - f(x) \leq \frac{f(x + t(y-x)) - f(x)}{t} \quad \Leftrightarrow \quad f(y) - f(x) \leq \frac{f(x+h) - f(x)}{h}(y - x),$$

for $h = t(y - x)$. Taking limits when $h \to 0$, the latter implies that

$$f(y) - f(x) \leq f'(x)(y - x). \tag{1}$$

To prove the converse, let $z = ty + (1-t)x$. By applying equation (1), we have that

$$f(y) - f(z) \leq f'(z)(y - z),$$
$$f(x) - f(z) \leq f'(z)(x - z).$$

Summing the first inequality multiplied by t to the second multiplied by $(1-t)$, we find the result

$$[f(y) - f(z)]t + [f(x) - f(z)](1-t) \leq 0 \quad \Leftrightarrow \quad tf(y) + (1-t)f(x) \leq f(z).$$

Claim: A twice differentiable function f on an open convex subset U of $\mathbb{R}^n$ is concave on U if and only if its Hessian $D^2 f(\mathbf{x})$ is negative semidefinite for all $\mathbf{x}$ in U. The function f is a convex function if and only if $D^2 f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x}$ in U.

Proof: The function $f : \mathbb{R} \to \mathbb{R}$ is concave if and only if

$$f(y) - f(x) \leq f'(x)(y - x) \;\text{ and }\; f(x) - f(y) \leq f'(y)(x - y).$$

Summing these and dividing by $-(y-x)^2$ implies that

$$\frac{f'(y) - f'(x)}{y - x} \leq 0.$$

The limit of the last expression when $y \to x$ yields $f''(x) \leq 0$. To prove the converse, pick $y \geq x$, and note that by Fundamental Theorem of Calculus and $f'' \leq 0$, we have that

$$f(y) - f(x) = \textstyle\int_x^y f'(t)dt \leq \int_x^y f'(x)dt = f'(x)(y - x).$$

Claim: Let $f_1, ..., f_k$ be concave functions defined on the same convex set $U \subset \mathbb{R}^n$. Let $a_1, a_2, ..., a_k$ be positive numbers. If so, the function

$$g = a_1 f_1 + a_2 f_2 + ... + a_k f_k$$

is a concave function on U.

Proof: Immediately observe:

$$\begin{aligned} g(t\mathbf{x} + (1-t)\mathbf{y}) &= \textstyle\sum_{i=1}^{k} a_i f_i(t\mathbf{x} + (1-t)\mathbf{y}) \geq \sum_{i=1}^{k} a_i (t f_i(\mathbf{x}) + (1-t) f_i(\mathbf{y})) = \\ &= t \textstyle\sum_{i=1}^{k} a_i f_i(\mathbf{x}) + (1-t) \sum_{i=1}^{k} a_i f_i(\mathbf{y}) = t g(\mathbf{x}) + (1-t) g(\mathbf{y}). \end{aligned}$$

Example: Consider a firm whose production function is $y = g(\mathbf{x})$, where y denotes output and $\mathbf{x}$ denote the input bundle. If p denotes the price of output and w_i is the cost per unit of input i, then the firm's profit function satisfies

$$\Pi(\mathbf{x}) = p g(\mathbf{x}) - (w_1 x_1 + w_2 x_2 + ... + w_n x_n).$$

Thus, the profit function is a concave function if the production function g is concave, since the previous claim holds and because $-(w_1 x_1 + w_2 x_2 + ... + w_n x_n)$ is linear and, thus concave.

Definition: Consider a real valued function f defined on $U \subset \mathbb{R}^n$:
(1) its *a-level set* is

$$C_a = \{\mathbf{x} \in U | f(\mathbf{x}) = a\};$$

(2) its *a-uppercontour* set is

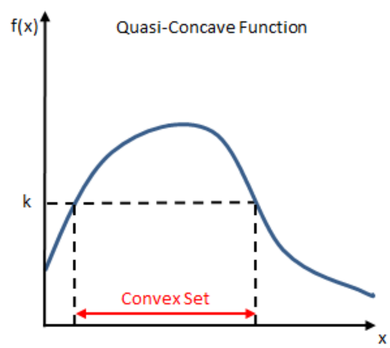
$$C_a^+ = \{\mathbf{x} \in U | f(\mathbf{x}) \geq a\};$$

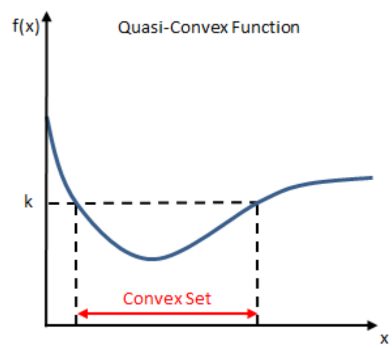
(3) its *a-lowercontour* set is

$$C_a^- = \{\mathbf{x} \in U | f(\mathbf{x}) \leq a\}.$$

Definition: A real valued function f defined on a convex set $U \subset \mathbb{R}^n$ is:
(1) *quasiconcave* if C_a^+ is a convex set for every real number a;
(2) *quasiconvex* if C_a^- is a convex set for every real number a.

The level sets of a quasiconcave function bound convex subsets from below.
The level sets of a quasiconvex function bound convex subsets from above.

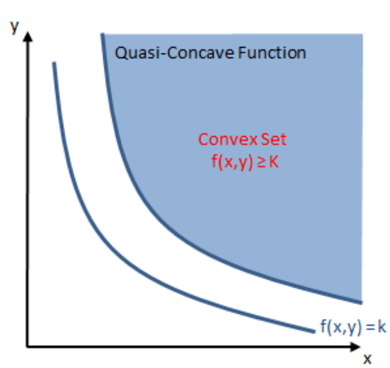




Claim: Every concave function is quasiconcave and every convex function is quasiconvex.
Proof: Let $\mathbf{x}, \mathbf{y} \in C_a^+$, so that $f(\mathbf{x}) \geq a$ and $f(\mathbf{y}) \geq a$, and let f be concave. If so:

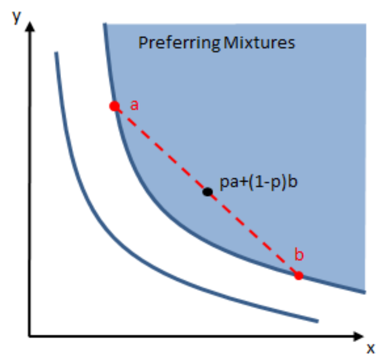
$$\begin{aligned} f(t\mathbf{x} + (\mathbf{1} - \mathbf{t})\mathbf{y}) &\geq t f(\mathbf{x}) + (1-t) f(\mathbf{y}) \\ &\geq ta + (1-t)a = a. \end{aligned}$$

So $t\mathbf{x} + (\mathbf{1} - \mathbf{t})\mathbf{y}$ is in C_a^+ and hence this set is convex. We have shown that if f is concave, it is also quasiconcave. Try to show that every convex function is quasiconvex.

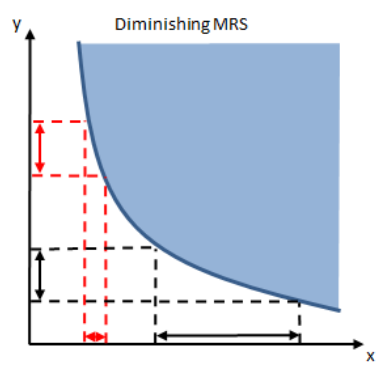


Quasi-concavity is simply a desirable property when we talk about economic objective functions such as preferences:

(1) The convexity of uppercontour sets is a natural requirement for utility and production functions. For example, consider an indifference curve of the concave utility function $f(x, y)$. Take two bundles on this indifference curve. The set of bundles which are preferred to them, is a convex set. In particular, the bundles that mix their contents are in this preferred set. Thus, a consumer with a concave utility function always prefers a mixture of any two bundles with the same utility to any of them.



(2) A more important advantage of the shape of the indifference curve is that it displays a diminishing marginal rate of substitution. As one moves left to right along the indifference curve increasing consumption of good x, the consumer is willing to give up more and more units of good x to gain an additional unit of good y. This property is a property of concave utility functions because each level set forms the boundary of a convex region.



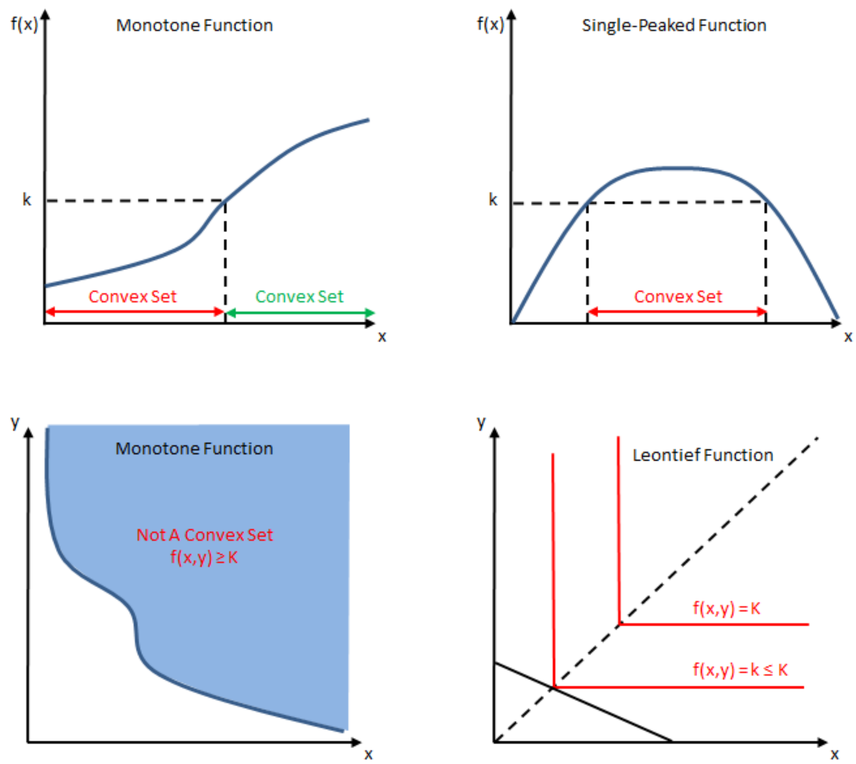
Observations:

(1) Any increasing transformation of a concave function is quasiconcave.

(2) An monotone function $f(x)$ on $\mathbb{R}^1$ is both quasiconcave and quasiconvex.

(3) A single peaked function $f(x)$ on $\mathbb{R}^1$ is quasiconcave.

(4) The function $\min\{x, y\}$ is quasiconcave, as C_a^+ is convex.



Proof of (1): Let f be quasiconcave and defined on a convex set U in $\mathbb{R}^n$. Consider an increasing transformation $g : \mathbb{R} \to \mathbb{R}$ with $g' > 0$ (you will prove the case in which $g' \geq 0$ in the problem set). For any function h, let $C_k^+(h)$ denote its uppercontour sets. So if $h(\mathbf{x}) = g(f(\mathbf{x}))$, we have that

$$C_k^+(f) = C_{g(k)}^+(h).$$

But as $C_k^+(f)$ is convex for all k, so is $C_{g(k)}^+(h)$; and hence h is quasi-concave.

Example: To show why monotone functions $\mathbb{R}^2_+$ are not quasiconcave, consider the function

$$f(x,y) = x^2 + y^2.$$

Such function is clearly increasing in $\mathbb{R}^2_+$ as $Df(x,y) = (2x, 2y)$. However, the function is not quasi-concave. In fact, $(1,0), (0,1) \in C_1^+$. However, $(1/2, 1/2) \notin C_1^+$ as

$$f(1/2, 1/2) = 1/2 < 1.$$

Claim: Let f be a function defined on a convex set $U \subset \mathbb{R}^n$. The following are equivalent:

1. f is a quasiconcave function on U;
2. for all $\mathbf{x}, \mathbf{y} \in U$ and $t \in [0,1]$,

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y})\}.$$

Proof: To prove that (1) implies (2), take any $\mathbf{x}, \mathbf{y} \in U$ and let $\min\{f(\mathbf{x}), f(\mathbf{y})\} = k$. The latter implies that $\mathbf{x}, \mathbf{y} \in C_k^+$. But if f is quasiconcave, C_k^+ must be convex and thus $t\mathbf{x}+(1-t)\mathbf{y} \in C_k^+$ for all $t \in [0,1]$. Thus, for any $t \in [0,1]$,

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \geq k = \min\{f(\mathbf{x}), f(\mathbf{y})\}.$$

To prove the converse, take any $\mathbf{x}, \mathbf{y} \in C_k^+$ and note that (2) implies that for all $t \in [0,1]$

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y}) \geq k.$$

Thus, $t\mathbf{x} + (1-t)\mathbf{y} \in C_k^+$ for all $t \in [0,1]$ and all $\mathbf{x}, \mathbf{y} \in C_k^+$, meaning that C_k^+ is convex for all k, and hence f is quasiconcave.

Lecture 2: Optimization

Definition of Extreme Points

Optimization plays a crucial role in economic problems. We start by analyzing unconstrained optimization problems.

Definition: The *ball* $B(\mathbf{x}|\varepsilon)$ centred at $\mathbf{x}$ of radius ε is the set of all vectors $\mathbf{y}$ in $\mathbb{R}^n$ whose distance from $\mathbf{x}$ is less than ε,

$$B(\mathbf{x}|\varepsilon) = \{\mathbf{y} \in \mathbb{R}^n | \varepsilon > ||\mathbf{y} - \mathbf{x}||\}.$$

Definition: Let $f(\mathbf{x})$ be a real valued function defined on $C \subseteq \mathbb{R}^n$. A point $\mathbf{x}^* \in C$ is:
(1) A *global maximizer* for $f(\mathbf{x})$ on C if

$$f(\mathbf{x}^*) \geq f(\mathbf{x}) \text{ for all } \mathbf{x} \in C.$$

(2) A *strict global maximizer* for $f(\mathbf{x})$ on C if

$$f(\mathbf{x}^*) > f(\mathbf{x}) \text{ for all } \mathbf{x} \in C \text{ such that } \mathbf{x} \neq \mathbf{x}^*.$$

(3) A *local maximizer* for $f(\mathbf{x})$ if there is a strictly positive number ε such that:

$$f(\mathbf{x}^*) \geq f(\mathbf{x}) \text{ for all } \mathbf{x} \in C \cap B(\mathbf{x}^*, \varepsilon).$$

(4) A *strict local maximizer* for $f(\mathbf{x})$ if there is a strictly positive number ε such that:

$$f(\mathbf{x}^*) > f(\mathbf{x}) \text{ for all } \mathbf{x} \in C \cap B(\mathbf{x}^*, \varepsilon) \text{ such that } \mathbf{x} \neq \mathbf{x}^*.$$

(5) A *critical point* for $f(\mathbf{x})$ if the first partial derivative of $f(\mathbf{x})$ exists at $\mathbf{x}^*$ and

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0 \text{ for } i = 1, 2, ..., n \quad \Leftrightarrow \quad Df(\mathbf{x}^*) = \mathbf{0}.$$

Example: To find the critical points of $F(x, y) = x^3 - y^3 + 9xy$, set

$$\frac{\partial F}{\partial x}(x, y) = 3x^2 + 9y = 0 \text{ and } \frac{\partial F}{\partial y}(x, y) = -3y^2 + 9x = 0.$$

The critical points are $(0, 0)$ and $(3, -3)$.

Do Extreme Points Exist?

Definition: A set is *compact* if it is closed and bounded.

Extreme Value Theorem: Let C be a compact subset of $\mathbb{R}^n$ and $f(x)$ be a continuous function defined on C. If so, there exists a point $\mathbf{x}^*$ in C which is a global maximizer of f, and a point $\mathbf{x}_*$ in C which is a global minimizer of f. That is,

$$f(\mathbf{x}_*) \leq f(\mathbf{x}) \leq f(\mathbf{x}^*) \quad \text{for all} \quad \mathbf{x} \in C.$$

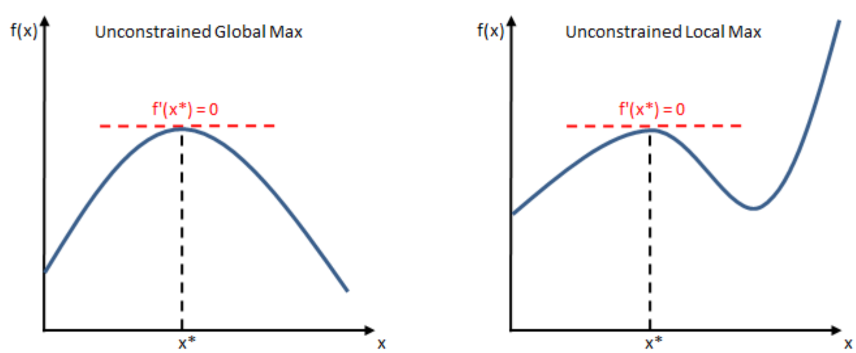
Unconstrained Optimization: Functions of One Variable

Consider real valued functions $f(x)$ that are continuously differentiable function on an interval I (so that, $f'(x)$ exists and is continuous).

First order **necessary condition for maximum** in $\mathbb{R}$:
If x^* is a local maximizer of $f(x)$, then either x^* is an end point of I or $f'(x^*) = 0$.

Second order **sufficient condition for a maximum** in $\mathbb{R}$:
If $f''(x)$ is continuous on I and x^* is a critical point of $f(x)$, then x^* is:

- a global maximizer of f on I if $f''(x) \leq 0$ for all $x \in I$;
- a strict global maximizer of f on I if $f''(x) < 0$ for all $x \in I$ and $x \neq x^*$;
- a strict local maximizer of f on I if $f''(x^*) < 0$.



Consider real valued functions $f(\mathbf{x})$ that are continuously differentiable on a subset $C \subset \mathbb{R}^n$ (so that, all first partial derivatives exist and are continuous).

First order **necessary conditions for a maximum** in $\mathbb{R}^n$:
If $\mathbf{x}^*$ is in the interior of C and is a local maximizer of $f(\mathbf{x})$, then $\mathbf{x}^*$ is a critical point of f,

$$Df(\mathbf{x}^*) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}^*), ..., \frac{\partial f}{\partial x_n}(\mathbf{x}^*)\right) = \mathbf{0}.$$

How does one determine whether a critical point is a local maximum or a local minimum then? To this end one has to consider the Hessian of the map f (the matrix of the second order partial derivatives). By the Clairaut-Schwarz theorem, the Hessian is always a symmetric matrix as cross-partial derivatives are equal (if the function has continuous second order partial derivatives).

Second order **sufficient conditions for a local maximum** in $\mathbb{R}^n$:
Let $f(\mathbf{x})$ be a function for which all second partial derivatives exist and are continuous on a subset $C \subset \mathbb{R}^n$. If $\mathbf{x}^*$ is critical point of f and if $D^2 f(\mathbf{x}^*)$ is negative (positive) **definite**, then $\mathbf{x}^*$ is a strict local maximizer (minimizer) of f.

Observations:
(1) If $\mathbf{x}^*$ is an interior point and a global maximum (minimum) of f, then $D^2 f(\mathbf{x}^*)$ is negative (positive) **semidefinite**.
(2) If $\mathbf{x}^*$ is a critical point, and $D^2 f(\mathbf{x}^*)$ is negative (positive) **semidefinite**, then $\mathbf{x}^*$ may not be local maximum (minimum).

For a counterexample to (2) consider $f(x) = x^3$. Clearly, $Df(0) = 0$ and $D^2 f(0) = 0$ is semidefinite. But $x = 0$ is neither a maximum or minimum.

Example: Again consider the example $f(x_1, x_2) = x_1^3 - x_2^3 + 9x_1x_2$. Compute the Hessian:

$$D^2 f(x_1, x_2) = \begin{pmatrix} 6x_1 & 9 \\ 9 & -6x_2 \end{pmatrix}$$

The first order leading principle minor is $6x_1$ and the second order leading principal minor is $\det(D^2 f(x_1, x_2)) = -36x_1x_2 - 81$. At $(0,0)$ the two minors are 0 and -81 and hence the matrix is indefinite. The point is neither a local min nor a local max (it is a saddle point). At $(3,-3)$ the two minors are positive and hence the point is a strict local minimum of f. However, this is not a global minimum (why? set $x_1 = 0$ and $x_2 \to \infty$).

Proof of Local SOC: By Taylor's Theorem we have that

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + Df(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^*)\mathbf{h} + R_2(\mathbf{h}|\mathbf{x}^*).$$

Ignore $R_2(\mathbf{h}|\mathbf{x}^*)$ as it is small and note that by assumption $Df(\mathbf{x}^*) = \mathbf{0}$. If so,

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^*)\mathbf{h}.$$

If $D^2 f(\mathbf{x}^*)$ is negative definite, for all small enough $\mathbf{h} \neq \mathbf{0}$**,** the right hand side is negative. This implies that

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) < 0,$$

or in other words $\mathbf{x}^*$ is a strict local maximizer of f.

Claim: If f is a differentiable concave function defined on a convex set U and if $\mathbf{x}, \mathbf{y} \in U$, then

$$Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq 0 \quad \Rightarrow \quad f(\mathbf{y}) \leq f(\mathbf{x}).$$

Moreover, $\mathbf{x}^*$ is a global maximizer of f if one of the following holds: (i) $Df(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \leq 0$ for all $\mathbf{y} \in U$; (ii) $Df(\mathbf{x}^*) = 0$.

Proof: As the function is differentiable, concavity is equivalent to

$$f(\mathbf{y}) - f(\mathbf{x}) \leq Df(\mathbf{x})(\mathbf{y} - \mathbf{x}).$$

If so, since by assumption $Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq 0$ we must have that

$$f(\mathbf{y}) - f(\mathbf{x}) \leq 0.$$

If follows then immediately that fulfilling (i) or (ii) implies that $\mathbf{x}^*$ is a global maximizer of f.

Second order **sufficient conditions for global maximum** (minimum) in $\mathbb{R}^n$:

Let $\mathbf{x}^*$ be a critical point of a twice differentiable function $f(\mathbf{x})$ on $\mathbb{R}^n$. Then, $\mathbf{x}^*$ is:

(1) a global maximizer for f if $D^2 f(\mathbf{x})$ is negative (positive) semidefinite on $\mathbb{R}^n$;

(2) a strict global maximizer for f if $D^2 f(\mathbf{x})$ is negative (positive) definite on $\mathbb{R}^n$.

Comment: The property that critical points of concave functions are global maximizers is an important one in economic theory. For example, many economic principles, such as marginal rate of substitution equals the price ratio, or marginal revenue equals marginal cost are simply the first order necessary conditions of the corresponding maximization problem. Ideally, one would like such conditions also to be a sufficient for guaranteeing that utility or profit is being maximized. This will indeed be the case when the objective function is concave.

Consider a subset $U \subseteq \mathbb{R}^n$, a vector $x^* \in U$ and a twice continuously differentiable objective function $f : U \to \mathbb{R}$.

Recall that $B(\mathbf{x}|\varepsilon)$ denotes an open ball of radius epsilon around $\mathbf{x}$, and let $\bar{B}(\mathbf{x}|\varepsilon) = B(\mathbf{x}|\varepsilon)\backslash\{\mathbf{x}\}$.

Consider the following conditions:

(A) $Df(\mathbf{x}^*) = 0$ & $D^2f(\mathbf{x}^*) \leq 0$.

(B) $Df(\mathbf{x}^*) = 0$, $D^2f(\mathbf{x}^*) \leq 0$, & $D^2f(\mathbf{x}) < 0$ for any $\mathbf{x} \in \bar{B}(\mathbf{x}^*|\varepsilon)$ for some $\varepsilon > 0$.

(C) $Df(\mathbf{x}^*) = 0$ & $D^2f(\mathbf{x}) \leq 0$ for any $\mathbf{x} \in U$.

(D) $Df(\mathbf{x}^*) = 0$ & $D^2f(\mathbf{x}) \leq 0$ for any $\mathbf{x} \in B(\mathbf{x}^*|\varepsilon)$ for some $\varepsilon > 0$.

(E) $Df(\mathbf{x}^*) = 0$, $D^2f(\mathbf{x}) \leq 0$ for any $\mathbf{x} \in U$ & $D^2f(\mathbf{x}) < 0$ for any $\mathbf{x} \in \bar{B}(\mathbf{x}^*|\varepsilon)$ for some $\varepsilon > 0$.

The following table summarizes necessary and sufficient conditions for a maximum:

	Necessary	Sufficient
Strict Global Max	B	E
Global Max	A or D	C
Strict Local Max	B	B
Local Max	A or D	D

We now analyze an optimization problem in which constraints limit the choice of variable

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \ \text{ subject to}$$
$$g_1(\mathbf{x}) \le 0, \ldots , g_k(\mathbf{x}) \le 0,$$
$$h_1(\mathbf{x}) = 0, \ldots , h_m(\mathbf{x}) = 0.$$

The function f is called *objective function.*

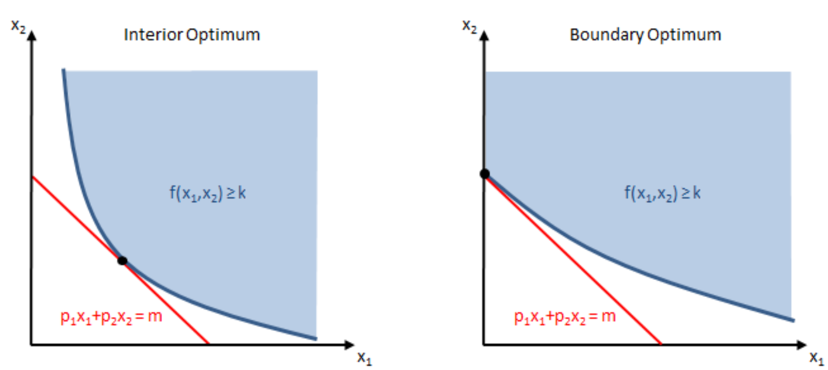
The functions g are called *inequality constraints.*

The functions h are called *equality constraints.*

Example: The classical utility maximization problem:

$$\max_{\mathbf{x}\in\mathbb{R}^n} U(x_1, x_2, \ldots, x_n) \ \text{ subject to}$$
$$p_1x_1 + p_2x_2 + \ldots + p_nx_n \le m$$
$$x_1 \ge 0,\ x_2 \ge 0, \ldots , x_n \ge 0$$

where we can treat the latter constraints as $-x_i \le 0$.



Consider a problem with two variables and one equality constraint

$$\max_{x_1,x_2} f(x_1,x_2) \text{ subject to}$$
$$p_1x_1 + p_2x_2 = m$$

Geometrical Representation:
Draw the constraint on the (x_1, x_2) plane.
Draw representative samples of level curves of the objective function f.
The goal is to find the highest value level curve of f which meets the constraint.
It cannot cross the constraint set; it therefore must be tangent to it.

To find the slope of the level set $f(x_1,x_2) = k$, use total differentiation:

$$\frac{\partial f(x_1,x_2)}{\partial x_1}dx_1 + \frac{\partial f(x_1,x_2)}{\partial x_2}dx_2 = 0 \quad \Rightarrow \quad \frac{dx_2}{dx_1} = -\frac{\partial f(x_1,x_2)}{\partial x_1} \Big/ \frac{\partial f(x_1,x_2)}{\partial x_2}.$$

The slope of the constraint similarly amounts to

$$\frac{dx_2}{dx_1} = -\frac{\partial h(x_1,x_2)}{\partial x_1} \Big/ \frac{\partial h(x_1,x_2)}{\partial x_2}.$$

Hence, tangency at the optimal $\mathbf{x}^*$ requires

$$\frac{\frac{\partial f}{\partial x_1}(\mathbf{x}^*)}{\frac{\partial f}{\partial x_2}(\mathbf{x}^*)} = \frac{\frac{\partial h}{\partial x_1}(\mathbf{x}^*)}{\frac{\partial h}{\partial x_2}(\mathbf{x}^*)}.$$

Manipulate the previous equation and let μ be common value satisfying

$$\mu \equiv \frac{\frac{\partial f}{\partial x_1}(\mathbf{x}^*)}{\frac{\partial h}{\partial x_1}(\mathbf{x}^*)} = \frac{\frac{\partial f}{\partial x_2}(\mathbf{x}^*)}{\frac{\partial h}{\partial x_2}(\mathbf{x}^*)}$$

If so, it is possible to re-write these two equations as:

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) - \mu\frac{\partial h}{\partial x_1}(\mathbf{x}^*) = 0$$
$$\frac{\partial f}{\partial x_2}(\mathbf{x}^*) - \mu\frac{\partial h}{\partial x_2}(\mathbf{x}^*) = 0$$

The solution of the maximization problem is thus defined the following system of three equations in three unknowns

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) - \mu\frac{\partial h}{\partial x_1}(\mathbf{x}^*) = 0$$
$$\frac{\partial f}{\partial x_2}(\mathbf{x}^*) - \mu\frac{\partial h}{\partial x_2}(\mathbf{x}^*) = 0$$
$$h(x_1^*, x_2^*) - c = 0$$

Alternatively, form the *Lagrangian function* associated to the maximization problem:

$$L(x_1, x_2, \mu) = f(x_1, x_2) - \mu(h(x_1, x_2) - c)$$

and find the critical point of L, by solving:

$$\frac{\partial L}{\partial x_1} = 0, \ \frac{\partial L}{\partial x_2} = 0 \ \text{ and } \ \frac{\partial L}{\partial \mu} = 0$$

and note that this results in the same system equations defining the optimum. The variable μ is called the *Lagrange multiplier.*

The Lagrange approach allowed us to reduce a constrained optimization problem in two variables to an unconstrained problem in three variables.

Caveat: The procedure fails if $\frac{\partial h}{\partial x_1}(\mathbf{x}^*) = \frac{\partial h}{\partial x_2}(\mathbf{x}^*) = 0$.
A **constraint qualification** must be added requiring $\mathbf{x}^*$ **not to be a critical point of** h, or formally $Dh(\mathbf{x}^*) \neq \mathbf{0}$. The procedure is not well defined when a critical point of the constraint belongs to the constraint set since the slope of the constraint is not defined at that point.

Lagrange Theorem: Let f and h be continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. Let $\mathbf{x}^*$ maximize $f(\mathbf{x})$ subject to $h(\mathbf{x}) = 0$ and suppose that $\mathbf{x}^*$ is not a critical point of h. If so, there is a real number μ^* such that $(\mathbf{x}^*, \mu^*)$ is a critical point of the Lagrange function

$$L(\mathbf{x}, \mu) = f(\mathbf{x}) - \mu h(\mathbf{x}).$$

Intuition: The Lagrangian amounts to the objective function diminished by a penalty (the Lagrange multiplier) for violating the constraint. As the constraint holds with equality when $\partial L/\partial \mu = 0$, no penalty is ever paid at the optimum. But the Lagrange multiplier characterizes the smallest fine for which it is not worthwhile to violate the constraint. The same analysis easily extends to the case of several equality constraints.

Example: Consider the following problem:

$$\max_{x_1, x_2} \ x_1 x_2 \ \text{ subject to } x_1 + 4x_2 = 16.$$

The constraint qualification is satisfied trivially, the Lagrangian amounts to

$$L(x_1, x_2, \mu) = x_1 x_2 - \mu(x_1 + 4x_2 - 16),$$

and the first order conditions amount to

$$\begin{aligned} x_2 - \mu &= 0, \\ x_1 - 4\mu &= 0, \\ x_1 + 4x_2 - 16 &= 0. \end{aligned}$$

The only solution requires $x_1 = 8$, $x_2 = 2$, $\mu = 2$.

Lagrange Theorem: Let f and h_i for all $i \in \{1, ..., m\}$ be continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. Let $\mathbf{x}^*$ maximize $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and suppose that the gradients $Dh_1(\mathbf{x}^*), ..., Dh_m(\mathbf{x}^*)$ are linearly independent. If so, there exists a vector $\boldsymbol{\mu}^* = (\mu_1^*, ..., \mu_m^*)$ such that

$$\underbrace{Df(\mathbf{x}^*)}_{1\times n} = \underbrace{\boldsymbol{\mu}^*}_{1\times m}\underbrace{D\mathbf{h}(\mathbf{x}^*)}_{m\times n} = \textstyle\sum_{i=1}^m \mu_i^* Dh_i(\mathbf{x}^*).$$

Proof: For convenience, consider the norm $\|\mathbf{h}\| = (\mathbf{h}^T\mathbf{h})^{1/2}$. For $a, c > 0$, define the following auxiliary function

$$f^a(\mathbf{x}) = f(\mathbf{x}) - \frac{a}{2}\|\mathbf{h}(\mathbf{x})\|^2 - \frac{c}{2}\|\mathbf{x} - \mathbf{x}^*\|^2.$$

Consider the neighborhood $B(\mathbf{x}^*|\varepsilon) = \{\mathbf{x} \in \mathbb{R}^n \mid \varepsilon \geq \|\mathbf{x} - \mathbf{x}^*\|\}$. Let $\mathbf{x}^a$ maximize $f^a(\mathbf{x})$ subject to $\mathbf{x} \in B(\mathbf{x}^*|\varepsilon)$ (where $\mathbf{x}^a$ exists by the extreme value theorem). For any $a > 0$, by optimality of $\mathbf{x}^a$ for all $\mathbf{x} \in B(\mathbf{x}^*|\varepsilon)$,

$$f^a(\mathbf{x}^a) \geq f^a(\mathbf{x}^*) = f(\mathbf{x}^*). \tag{2}$$

As $f^a(\mathbf{x})$ is bounded on B_ε^* by continuity, we have that

$$\lim_{a\to\infty}\|\mathbf{h}(\mathbf{x}^a)\|^2 = 0,$$

or else $\lim_{a\to\infty} f^a(\mathbf{x}^a) = -\infty$. Thus, if $\lim_{a\to\infty}\mathbf{x}^a = \bar{\mathbf{x}}$, we must have that $\|\mathbf{h}(\bar{\mathbf{x}})\|^2 = 0$. Taking limits of equation (2) implies that

$$\lim_{a\to\infty} f^a(\mathbf{x}^a) = f(\bar{\mathbf{x}}) - \frac{c}{2}\|\bar{\mathbf{x}} - \mathbf{x}^*\|^2 \geq f(\mathbf{x}^*).$$

But if so, $\bar{\mathbf{x}} = \mathbf{x}^*$, given that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x}^*)$ as $\mathbf{x}^*$ maximizes $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$.

Necessary conditions for the optimality of $\mathbf{x}^a$ require

$$Df^a(\mathbf{x}^a) = \underbrace{Df(\mathbf{x}^a)}_{1\times n} - a\underbrace{\mathbf{h}(\mathbf{x}^a)^T}_{1\times m}\underbrace{D\mathbf{h}(\mathbf{x}^a)}_{m\times n} - c\underbrace{(\mathbf{x}^a - \mathbf{x}^*)^T}_{1\times n} = 0. \tag{3}$$

As $D\mathbf{h}(\mathbf{x}^*)$ is linearly independent, $D\mathbf{h}(\mathbf{x})$ must be linearly independent for any $\mathbf{x} \in B(\mathbf{x}^*|\varepsilon)$ when ε sufficiently small, since $\mathbf{h}$ is continuously differentiable. But if so, $\left[D\mathbf{h}(\mathbf{x})D\mathbf{h}(\mathbf{x})^T\right]^{-1}$ exists for all $\mathbf{x} \in B(\mathbf{x}^*|\varepsilon)$. Then, post-multiplying equation (3) by $D\mathbf{h}(\mathbf{x}^a)^T\left[D\mathbf{h}(\mathbf{x}^a)D\mathbf{h}(\mathbf{x}^a)^T\right]^{-1}$ implies that

$$(Df(\mathbf{x}^a) - c(\mathbf{x}^a - \mathbf{x}^*)^T)D\mathbf{h}(\mathbf{x}^a)^T\left[D\mathbf{h}(\mathbf{x}^a)D\mathbf{h}(\mathbf{x}^a)^T\right]^{-1} = a\mathbf{h}(\mathbf{x}^a)^T.$$

Taking limits as a diverges, implies that

$$\lim_{a\to\infty} a\mathbf{h}(\mathbf{x}^a)^T = Df(\mathbf{x}^*)D\mathbf{h}(\mathbf{x}^*)^T\left[D\mathbf{h}(\mathbf{x}^*)D\mathbf{h}(\mathbf{x}^*)^T\right]^{-1} \equiv \boldsymbol{\mu}^*.$$

Taking limits of equation (3) then delivers the desired result since

$$Df(\mathbf{x}^*) = \boldsymbol{\mu}^* D\mathbf{h}(\mathbf{x}^*).$$

With equality constraints, the solution to a maximization problem satisfied

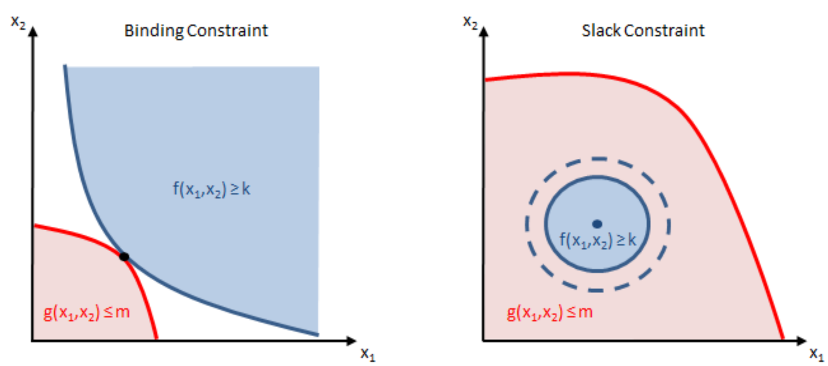
$$Df(\mathbf{x}^*) = \mu Dh(\mathbf{x}^*),$$

and no further restriction necessary on μ.

Now consider a problem with two variables and one inequality constraint,

$$\max_{x_1,x_2} f(x_1,x_2) \text{ subject to } g(x_1,x_2) \leq 0.$$

Graphical Representation:
Either the solution is on the boundary of the constraint set.
If so, the constraint is *binding* and there is a tangency at the solution.
Or the solution is in the interior of the constraint set.
If so, the constraint is *slack* and the solution is unaffected.



When the constraint binds $g(x_1,x_2) = 0$, the problem is similar to one with an equality constraint. However, even when the constraint binds, a restriction has to be imposed on the sign of Lagrange multiplier μ. Gradients still satisfy

$$Df(\mathbf{x}^*) = \mu Dg(\mathbf{x}^*).$$

But now the sign of μ is important. In fact, the gradients must point in the same direction or else f could increase and still satisfy the constraint. This means that $\mu \geq 0$. This is the main difference between inequality and equality constraints.

To solve such a problem, we still construct the Lagrangian

$$L(x_1,x_2,\mu) = f(x_1,x_2) - \mu g(x_1,x_2),$$

and then find the critical point of L, by setting

$$\frac{\partial L}{\partial x_i} = \frac{\partial f(x_1,x_2)}{\partial x_i} - \mu \frac{\partial g(x_1,x_2)}{\partial x_i} = 0 \text{ for all } i \in \{1,2\}.$$

But what about $\partial L/\partial \mu$?

(1) If the maximum $\mathbf{x}^*$ is achieved when $g(\mathbf{x}^*) = 0$, then the problem is solved as for an equality constraint and $\partial L/\partial \mu = 0$.

(2) If the maximum $\mathbf{x}^*$ is achieved when $g(\mathbf{x}^*) < 0$, the constraint is not binding and the optimal solution is interior. If so, the point $\mathbf{x}^*$ is an **unconstrained** local maximum and satisfies

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \frac{\partial f}{\partial x_2}(\mathbf{x}^*) = 0,$$

and therefore can still use the Lagrangian, provided that we set $\mu = 0$!!

In other words, either the constraint is binding and $g(x_1, x_2) = 0$, or it does not bind and $\mu = 0$ as a slack constraint cannot affect the solution to a maximization problem. In short, the following *complementary slackness* condition has to hold

$$\mu \frac{\partial L}{\partial \mu} = \mu g(x_1, x_2) = 0.$$

Recall that the constrained optimization problem with two variables and one inequality constraints. Let f and g be continuous differentiable functions. Let $\mathbf{x}^* = (x_1^*, x_2^*)$ be a solution to $\max f(\mathbf{x})$ subject to $g(\mathbf{x}) \leq 0$. If $g(\mathbf{x}^*) = 0$, suppose that $\mathbf{x}^*$ is not a critical point of g. Then, there is a real number μ^* such that

$$D_{\mathbf{x}} L(\mathbf{x}^*, \mu^*) = \mathbf{0}, \quad \mu^* D_\mu L(\mathbf{x}^*, \mu^*) = \mathbf{0}, \quad \mu^* \geq 0 \text{ and } g(\mathbf{x}^*) \leq 0.$$

where the Lagrangian amounts to $L(\mathbf{x}, \mu) = f(\mathbf{x}) - \mu g(\mathbf{x})$.

Example One Constraint: Consider a profit maximizing firm that produces output y from input x according to the production function $x^{0.5}$. Let the price of output be 2, and that of input be 1. The firm cannot buy more than $a > 0$ units of input. This firm's maximization problem amounts to

$$\max_x 2x^{0.5} - x \text{ subject to } x - a \leq 0.$$

The Lagrangian associated to this problem is

$$L(x, \mu) = 2x^{0.5} - x - \mu[x - a].$$

First order necessary conditions for this problem require

$$x^{-0.5} - 1 = \mu, \;\; \mu(x - a) = 0, \;\; \mu \geq 0 \text{ and } x - a \leq 0.$$

To solve the system of equations consider two cases:
(1) If $\mu > 0$, the constraint must bind. Thus, $x = a$ and

$$\mu = f'(a) = a^{-0.5} - 1.$$

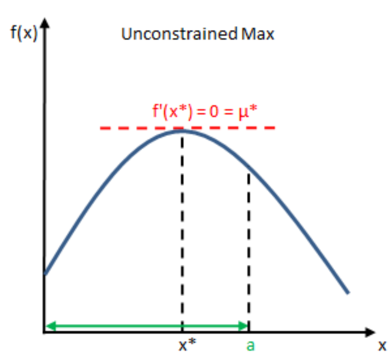
For this solution to be viable $\mu > 0$ which amounts to

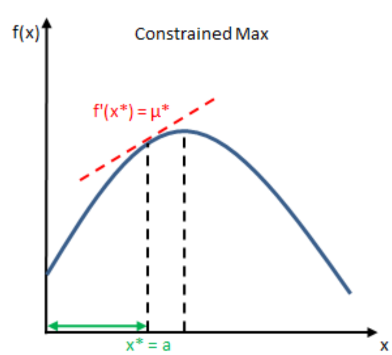
$$\mu = a^{-0.5} - 1 > 0 \Leftrightarrow a < 1.$$

(2) If $\mu = 0$, the constraint must be slack. From the first order condition,

$$x^{-0.5} - 1 = 0 \Leftrightarrow x = 1.$$

The solution thus requires $x = 1$ and $\mu = 0$, which is viable only if $a \geq 1$.





The generalization to several inequality constraints is straightforward. However, now a subset of constraints may be binding at the optimum.

Example More Constraints: Consider the following maximization problem:

$$\max x_1 x_2 x_3 \ \text{ subject to } \ x_1 + x_2 + x_3 \leq 3, \ \ x_1 \geq 0, \ \ x_2 \geq 0 \ \text{ and } \ x_3 \geq 0.$$

The Lagrangian associated to this problem satisfies:

$$L(\mathbf{x}, \boldsymbol{\mu}) = x_1 x_2 x_3 + \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3 - \mu_4(x_1 + x_2 + x_3 - 3).$$

Before solving the Lagrangian, observe that at the relevant critical points $\mu_1 = \mu_2 = \mu_3 = 0$.

In fact, if $\mu_i > 0$, complementary slackness would require $x_i = 0$ and $x_1 x_2 x_3 = 0$.

Obviously we could improve, for instance by setting $x_i = 1/2$ for all i.

What remains to be established is whether $\mu_4 > 0$ or $\mu_4 = 0$ at the optimum.

But, $\mu_4 > 0$ since if the constraint was slack

$$x_1 + x_2 + x_3 < 3,$$

and it would be possible to meet the constraint while increase the value of the function $x_1 x_2 x_3$ by increasing any one of the three variables.

Thus, from the first order conditions we obtain that

$$x_1 x_2 = x_2 x_3 = x_1 x_3 = \mu_4.$$

It follows that $\mu_4 = x_1 = x_2 = x_3 = 1$ at the solution.

Consider the following general constrained optimization problem

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0},$$

where $\mathbf{g}$ consists of m constraints $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), ..., g_m(\mathbf{x}))^T$. For this problem we have characterized necessary conditions for a maximum. These conditions state that any solution $\mathbf{x}^*$ of the constrained optimization problem must also be a critical point of the Lagrangian,

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \mu\mathbf{g}(\mathbf{x}),$$

where $\boldsymbol{\mu}^* = (\mu_1^*, .., \mu_m^*)$ denotes the vector of multipliers.

Problem: We would also like to know:

- whether any critical point of the Lagrangian is a maximizer of the Lagrangian;
- whether any maximizer of the Lagrangian, solves the constrained optimization problem.

Solution: To answer the latter, let $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ satisfy necessary conditions for a maximum:

$$\frac{\partial L(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial x_i} = \frac{\partial f(\mathbf{x}^*)}{\partial x_i} - \boldsymbol{\mu}^* \frac{\partial \mathbf{g}(\mathbf{x}^*)}{\partial x_i} = 0 \text{ for any } i \in \{1, ..., n\},$$
$$\frac{\partial L(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial \mu_j}\mu_j^* = 0, \ \ \mu_j^* \geq 0 \text{ and } g_j(\mathbf{x}^*) \leq 0 \text{ for any } j \in \{1, ..., m\}.$$

Claim: If $\mathbf{x}^*$ maximizes the Lagrangian, it also maximizes f subject to $\mathbf{g}(\mathbf{x}) = 0$.

Proof: To see this note that by complementary slackness $\boldsymbol{\mu}^*\mathbf{g}(\mathbf{x}^*) = 0$ and thus

$$L(\mathbf{x}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*) - \mu^*\mathbf{g}(\mathbf{x}^*) = f(\mathbf{x}^*).$$

By $\boldsymbol{\mu} \geq 0$ and $\mathbf{g}(\mathbf{x}) \leq 0$ for all other $\mathbf{x}$, we get that

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \mu\mathbf{g}(\mathbf{x}) \geq f(\mathbf{x}).$$

Moreover, since $\mathbf{x}^*$ maximizes the Lagrangian, for all $\mathbf{x}$,

$$L(\mathbf{x}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*) - \mu^*\mathbf{g}(\mathbf{x}^*) \geq f(\mathbf{x}) - \mu^*\mathbf{g}(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\mu}^*),$$

which implies that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}).$$

Thus, if $\mathbf{x}^*$ maximizes the Lagrangian, it also solves $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$.

To conclude the argument recall the main results from unconstrained optimization. If f is a concave function defined on a convex subset of $\mathbb{R}^n$ and $\mathbf{x}_0$ is a point in the interior in which $Df(\mathbf{x}_0) = 0$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for all $\mathbf{x}$. You have shown in class that in the constrained optimization problem, if f is concave and g is convex, then the Lagrangian function is also concave which means that first order conditions are sufficient in this case.

Again, consider the general constrained optimization problem

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ subject to } \mathbf{h}(\mathbf{x}) = \mathbf{0} \text{ and } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}.$$

We need to impose additional constraint qualifications to ensure that the problem is not degenerate. These can be stated in many ways. Two equivalent requirements appear below. Denote by $\mathbf{g}_E = (g_1, .., g_e)^T$ the vector of binding constraints (namely, those for which $(\mu_1, .., \mu_e) > \mathbf{0}$), while letting g_{e+1} to g_k denote slack ones. Posit that the multipliers associated to all equality constraints differ from 0 (or else, drop those for which this is not the case).

KKT Constraint Qualifications: Consider one of the following two restrictions:

- **Slater Condition**: there exists a point $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{g}(\mathbf{x}) < \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, and $\mathbf{h}$ is affine.
- **Linear Independence**: the gradients of the binding inequality constraints, $D_{\mathbf{x}}\mathbf{g}_E(\mathbf{x}^*)$, and the gradients of the equality constraints, $D_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*)$ are linearly independent at the solution $\mathbf{x}^*$.

Recall that linear independence amounts to requiring

$$\mathbf{w}^T \begin{pmatrix} D_{\mathbf{x}}\mathbf{g}_E(\mathbf{x}^*) \\ D_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*) \end{pmatrix} \neq \mathbf{0} \text{ for any } \mathbf{w} \in \mathbb{R}^{e+m} \backslash \{\mathbf{0}\}.$$

KKT Theorem: Let f, $\mathbf{h}$ and $\mathbf{g}$ be differentiable. Let f be concave, h_j be convex for all $j \in \{1, ..., k\}$, g_j be convex for all $j \in \{1, ..., m\}$, and a constraint qualification hold. Then $\mathbf{x}^*$ solves the constraint optimization problem if and only if $(\mathbf{x}^*, \lambda^*, \boldsymbol{\mu}^*)$ solves

$$Df(\mathbf{x}^*) - \lambda^* D\mathbf{h}(\mathbf{x}^*) - \boldsymbol{\mu}^* D\mathbf{g}(\mathbf{x}^*) = \mathbf{0},$$
$$\mathbf{h}(\mathbf{x}^*) = 0, \ \ \boldsymbol{\mu}^*\mathbf{g}(\mathbf{x}^*) = 0, \ \ \boldsymbol{\mu}^* \geq \mathbf{0} \text{ and } \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}.$$

Concavity of the objective function and convexity of the constraints imply that any critical point of the Lagrangian is a global solution of the constraint optimization problem. If such conditions are not satisfied, we can still rely on sufficient conditions for local maxima.

Mechanically, one can solve constrained optimization problems in the following way:

(1) Form the Lagrangian $L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \mu \mathbf{g}(\mathbf{x})$, and a solution to KKT necessary conditions:

$$D_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}, \;\; \boldsymbol{\mu}^* D_{\boldsymbol{\mu}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)^T = 0, \;\; \boldsymbol{\mu}^* \geq \mathbf{0} \;\text{ and }\; \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}.$$

(2) Let g_1 to g_e denote the binding constraints and g_{e+1} to g_k slack ones. Define $\mathbf{g}_E = (g_1, .., g_e)^T$.

(3) Observe that the set $\{\mathbf{v} \in \mathbb{R}^n | D_{\mathbf{x}} \mathbf{g}_E(\mathbf{x}^*)\mathbf{v} = 0\}$ identifies the hyperplane tangent to the constraint set at the critical point $\mathbf{x}^*$. For local second order conditions to hold we need the Lagrangian to be negative definite for any non-degenerate point in a neighborhood of $\mathbf{x}^*$ on such a linear subspace tangent to the constraint set.

(4) Formally this requires the Hessian of the Lagrangian with respect to $\mathbf{x}$ at the critical point $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ to be negative definite on the hyperplane $\{\mathbf{v} \in \mathbb{R}^n | D_{\mathbf{x}} \mathbf{g}_E(\mathbf{x}^*)\mathbf{v} = \mathbf{0}\}$. That is, for any $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{v} \neq \mathbf{0}$, we need that

$$D_{\mathbf{x}} \mathbf{g}_E(\mathbf{x}^*)\mathbf{v} = \mathbf{0} \quad \Rightarrow \quad \mathbf{v}^T (D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*))\mathbf{v} < 0,$$

If so, $\mathbf{x}^*$ is a strict local maximum of f on the constraint set.

(5) A simple way to check this condition can be obtained by forming the *bordered Hessian*,

$$D^2(\mathbf{x}^*, \boldsymbol{\mu}^*) = \begin{pmatrix} D^2_{\boldsymbol{\mu}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) & D_{\mu} D_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) \\ D_{\boldsymbol{\mu}} D_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)^T & D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) \end{pmatrix} = \begin{pmatrix} \mathbf{0} & D_{\mathbf{x}} \mathbf{g}_E(\mathbf{x}^*) \\ D_{\mathbf{x}} \mathbf{g}_E(\mathbf{x}^*)^T & D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) \end{pmatrix}.$$

(6) If the largest $n - e$ leading principal minors of $D^2(\mathbf{x}^*, \boldsymbol{\mu}^*)$ alternate in sign with the sign of the largest leading principal minor the same as the sign of $(-1)^n$, then *sufficient second order conditions* hold for the critical point to be a local maximum of the constrained maximization problem.

Remark: If the largest $n - e$ leading principal minors of $D^2(\mathbf{x}^*, \boldsymbol{\mu}^*)$ have the same sign as $(-1)^e$, then *sufficient second order conditions* hold for the critical point to be a local minimum of the constrained maximization problem.

The KKT Theorem applies to concave objective functions and convex constraints. Many economic objective functions are however only quasi-concave, and many constraints are only quasi-convex. The following result generalizes the KKT Theorem to these settings. For sake of brevity, we state the result only for inequality constraints. But the generalization is straightforward.

AE Theorem: Let f and $\mathbf{g}$ be differentiable. Let f be quasi-concave and g_j be quasi-convex for all j. Then, any solution $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ to the KKT necessary conditions,

$$D_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}, \;\; \boldsymbol{\mu}^* D_{\boldsymbol{\mu}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)^T = 0, \;\; \boldsymbol{\mu}^* \geq \mathbf{0} \;\text{ and }\; \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0},$$

solves the constraint optimization problem provided that $D_{\mathbf{x}} f(\mathbf{x}^*) \neq \mathbf{0}$ and that $f(\mathbf{x})$ is twice differentiable in the neighborhood of $\mathbf{x}^*$.

Numerous alternative regularity conditions can replace the latter requirement which only deals with slopes being well defined.

Lecture 3: Comparative Statics and Fixed Points

Value Functions and Envelope Theorem

The value function is the objective function of a constraint optimization problem evaluated at the solution. Profit functions and indirect utility functions are examples of maximum value functions, whereas cost functions and expenditure functions are minimum value functions.

Consider a general constrained optimization problem:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}|b) \text{ subject to } \mathbf{g}(\mathbf{x}|b) \leq \mathbf{0}.$$

where $\mathbf{g}(\mathbf{x}|b) = (g_1(\mathbf{x}|b), .., g_m(\mathbf{x}|b))^T$ denotes the vector of constraints and b a parameter.

For this constrained optimization problem:

- let $\mathbf{x}(b) = (x_1(b), ..., x_n(b))^T$ denote the optimal solution;
- let $\boldsymbol{\mu}(b) = (\mu_1(b), ..., \mu_m(b))$ denote the corresponding Lagrange multipliers;
- let $L(\mathbf{x}, \boldsymbol{\mu}|b)$ denote the Lagrangian of the constraint optimization problem;
- omit any slack constraints so that $\boldsymbol{\mu}(b) >> \mathbf{0}$.

Definition: The *maximum value function* of the problem amounts to

$$v(b) = f(\mathbf{x}(b)|b).$$

The Envelope Theorem relates changes in parameters to the value function.

Envelope Theorem: Suppose that $(\mathbf{x}(b), \boldsymbol{\mu}(b))$ are differentiable functions and that $\mathbf{x}(b)$ satisfies the constraint qualification. If so,

$$\frac{dv(b)}{db} = \frac{\partial L(\mathbf{x}(b), \boldsymbol{\mu}(b)|b)}{\partial b}.$$

Envelope Theorem: Suppose that $(\mathbf{x}(b), \boldsymbol{\mu}(b))$ are differentiable functions and that $\mathbf{x}(b)$ satisfies the constraint qualification. If so,

$$\frac{dv(b)}{db} = \frac{\partial L(\mathbf{x}(b), \boldsymbol{\mu}(b)|b)}{\partial b}.$$

Proof: Recall that FOC for the constraint optimization problem require that

$$\underbrace{D_{\mathbf{x}} f(\mathbf{x}(b)|b)}_{1\times n} \;=\; \underbrace{\boldsymbol{\mu}(b)}_{1\times m} \underbrace{D_{\mathbf{x}}\mathbf{g}(\mathbf{x}(b)|b)}_{m\times n}.$$

Observe that, as all the constraints bind, total differentiation implies that

$$\underbrace{D_{\mathbf{x}}\mathbf{g}(\mathbf{x}(b)|b)}_{m\times n} \underbrace{D_b\mathbf{x}(b)}_{n\times 1} \;+\; \underbrace{D_b\mathbf{g}(\mathbf{x}(b)|b)}_{m\times 1} \;= \mathbf{0}.$$

Finally note that by partially differentiating the Lagrangian we get

$$\underbrace{D_b L(\mathbf{x}(b), \boldsymbol{\mu}(b)|b)}_{1\times 1} \;=\; \underbrace{D_b f(\mathbf{x}(b)|b)}_{1\times 1} \;-\; \underbrace{\boldsymbol{\mu}(b)}_{1\times m} \underbrace{D_b\mathbf{g}(\mathbf{x}(b)|b)}_{m\times 1}.$$

To prove the result observe that

$$\begin{aligned}
\frac{dv(b)}{db} &= \frac{df(\mathbf{x}(b)|b)}{db} = D_b f(\mathbf{x}(b)|b) + D_{\mathbf{x}} f(\mathbf{x}(b)|b) D_b\mathbf{x}(b) = \\
&= D_b f(\mathbf{x}(b)|b) + \boldsymbol{\mu}(b) D_{\mathbf{x}}\mathbf{g}(\mathbf{x}(b)|b) D_b\mathbf{x}(b) = \\
&= D_b f(\mathbf{x}(b)|b) - \boldsymbol{\mu}(b) D_b\mathbf{g}(\mathbf{x}(b)|b) = \\
&= \frac{\partial L(\mathbf{x}(b), \boldsymbol{\mu}(b)|b)}{\partial b}.
\end{aligned}$$

where the first equality is definitional, the second holds by simple differentiation, the third holds by FOC, the fourth by totally differentiating constraints, and the fifth amounts to the partial of the Lagrangian.

The Meaning of Lagrange Multiplier

Consider a simple problem with n variables and 1 equality constraint

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) = b.$$

The Lagrangian of this problem satisfies

$$L(\mathbf{x}, \mu|b) = f(\mathbf{x}) - \mu(h(\mathbf{x}) - b).$$

For any value b, by FOC the solution satisfies:

$$D_{\mathbf{x}} f(\mathbf{x}(b)) - \mu(b) D_{\mathbf{x}} h(\mathbf{x}(b)) = 0.$$

Furthermore, since $h(\mathbf{x}(b)) = b$ for all b:

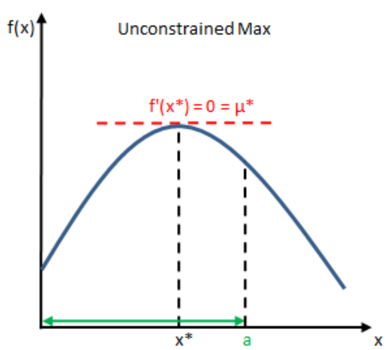
$$D_{\mathbf{x}} h(\mathbf{x}(b)) D_b \mathbf{x}(b) = 1$$

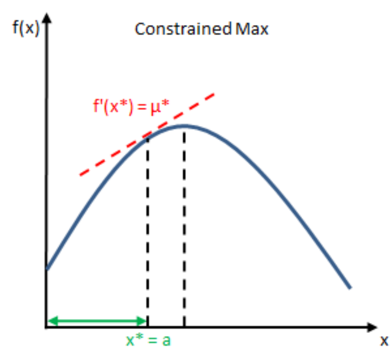
Therefore, using the chain rule we obtain the desired result

$$\frac{df(\mathbf{x}(b))}{db} = D_{\mathbf{x}} f(\mathbf{x}(b) D_b \mathbf{x}(b) = \mu(b) D_{\mathbf{x}} h(\mathbf{x}(b)) D_b \mathbf{x}(b) = \mu(b).$$

Interpretation: This observation allows us to attach to the multiplier the economic interpretation of a *shadow price*. For instance, in a model in which a firm maximizes profits subject to some resource constraint, the multiplier identifies how valuable an additional unit of input would be to the firm's profits, or how much the maximum value of the firm changes when the constraint is relaxed. In other words, it is the maximum amount the firm would be willing to pay to acquire another unit of input. This immediately follows as

$$\frac{d}{db} f(\mathbf{x}(b)) = \mu(b) = \frac{\partial}{\partial b} L(\mathbf{x}(b), \mu(b)|b).$$





Consider the implications of the envelope theorem for a scalar b in an unconstrained problem

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}|b).$$

Let $\mathbf{x}(b)$ denote the solution of this problem, and assume it to be continuous and differentiable in b.

In this setup the envelope theorem simplifies to

$$\frac{df(\mathbf{x}(b)|b)}{db} = \frac{\partial f(\mathbf{x}(b)|b)}{\partial \mathbf{x}}\frac{\partial \mathbf{x}(b)}{\partial b} + \frac{\partial f(\mathbf{x}(b)|b)}{\partial b} = \frac{\partial f(\mathbf{x}(b)|b)}{\partial b},$$

since by the first order conditions

$$\frac{\partial f(\mathbf{x}(b)|b)}{\partial \mathbf{x}} = \mathbf{0}.$$

Intuition: When we are already at a maximum, changing slightly the parameters of the problem or the constraints, does not affect the value through changes in the solution $\mathbf{x}(b)$, because the first order conditions hold. Thus, only the partial derivative matters. If there are multiple solutions and you can still use the envelope theorem, but you have to make sure that you do not jump to a different solution in a discrete manner.

Many relevant economic problems study how an equilibrium or a solution to an optimization problem is affected by changes in the exogenous variables. The implicit function theorem (IFT) is the prime tool for such comparisons.

The IFT provides sufficient conditions for a set of simultaneous equations

$$F_i(\mathbf{x}|\mathbf{b}) = F_i(x_1, ..., x_n|b_1, ..., b_m) = 0 \text{ for } \forall i = 1, ..., n \text{ or equivalently } \mathbf{F}(\mathbf{x}|\mathbf{b}) = \mathbf{0},$$

to have a solution defined by of implicit functions

$$x_i = x_i(b) = x_i(b_1, ..., b_m) \text{ for } \forall i = 1, ..., n \text{ or equivalently } \mathbf{x} = \mathbf{x}(\mathbf{b}).$$

In other words, IFT disciplines the system of equations so to ensure that the n equations defining the system can be solved by the n variables, $x_1, ..., x_n$, even when an explicit form solution cannot be obtained.

Implicit Function Theorem: Consider the system of equations $\mathbf{F}(\mathbf{x}|\mathbf{b}) = \mathbf{0}$. Let any function F_i have continuous partial derivatives with respect to all $\mathbf{x}$ and $\mathbf{b}$ variables. Consider a point $\bar{\mathbf{x}}$ that solves the system of equations at parameter values $\bar{\mathbf{b}}$ and at which the determinant of the $n \times n$ Jacobian with respect to the $\mathbf{x}$ variables is not zero,

$$\left|D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})\right| = \left|\frac{\partial \mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})}{\partial \mathbf{x}}\right| = \begin{vmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_n} \end{vmatrix} \neq 0.$$

If so, there exists an m-dimensional neighborhood of $\bar{\mathbf{b}}$ in which the variables $\mathbf{x}$ are functions of $\mathbf{b}$. For every vector $\mathbf{b}$ in the neighborhood $\mathbf{F}(\mathbf{x}(\mathbf{b})|\mathbf{b}) = \mathbf{0}$, thereby giving to the system of equations the status of a set of identities in this neighborhood. Moreover, the implicit functions $\mathbf{x}(\mathbf{b})$ are continuous and have continuous partial derivatives with respect to all the variables $\mathbf{b}$.

Thus, IFT allows us to find the partial derivatives of the implicit functions $\mathbf{x}(\mathbf{b})$ without solving for the $\mathbf{x}$ variables. To do so, since in the neighborhood of the solution the equations have a status of identities, we can set the total differential to zero, $dF(\mathbf{x}(\mathbf{b})|\mathbf{b})/d\mathbf{b} = 0$. For instance, when considering only $db_i \neq 0$ and setting $db_j = 0$ for $j \neq i$, the result in matrix notation implies that

$$\frac{\partial \mathbf{F}(\mathbf{x}(\mathbf{b})|\mathbf{b})}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(\mathbf{b})}{\partial b_i} + \frac{\partial \mathbf{F}(\mathbf{x}(\mathbf{b})|\mathbf{b})}{\partial b_i} = 0 \quad \Rightarrow \quad \frac{\partial \mathbf{x}(\mathbf{b})}{\partial b_i} = -\left[\frac{\partial \mathbf{F}(\mathbf{x}(\mathbf{b})|\mathbf{b})}{\partial \mathbf{x}}\right]^{-1} \frac{\partial \mathbf{F}(\mathbf{x}(\mathbf{b})|\mathbf{b})}{\partial b_i}.$$

Since $\left|D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})\right| \neq \mathbf{0}$, a unique nontrivial solution exists to this linear system which by Cramer's rule satisfies

$$\frac{\partial x_j(\mathbf{b})}{\partial b_i} = \frac{\left|D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})_j\right|}{\left|D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})\right|}$$

where $D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})_j$ is the matrix formed by replacing the j^{th} column of $D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})$ with the vector $-D_{b_i}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})$.

For $n = 2$ and $m = 1$, consider the system of equations

$$\mathbf{F}(\mathbf{x}|b) = \begin{cases} x_1 - e^{x_2} \\ x_1^2 + x_2^2 - b \end{cases} = \mathbf{0}.$$

Restrict attention to the case in which $\mathbf{x} \geq \mathbf{0}$. When $\mathbf{x} \neq \mathbf{0}$, we have that

$$|D_{\mathbf{x}}\mathbf{F}(\mathbf{x}|b)| = 2x_2 + 2x_1 e^{x_2} > 0.$$

Thus, when $\mathbf{x} \neq \mathbf{0}$, IFT applies and there exists implicit functions such that

$$\mathbf{F}(\mathbf{x}(b)|b) = \begin{cases} x_1(b) - e^{x_2(b)} \\ x_1(b)^2 + x_2(b)^2 - b \end{cases} = \mathbf{0}.$$

By totally differentiating the system, we find

$$\frac{d\mathbf{F}(\mathbf{x}(b)|b)}{db} = \begin{cases} x_1'(b) - x_2'(b)e^{x_2(b)} \\ 2x_1'(b)x_1(b) - 2x_2'(b)x_2(b) - 1 \end{cases} = \mathbf{0}.$$

Consequently, by solving the linear system, we find that

$$\frac{\partial \mathbf{x}(b)}{\partial b} = -\left[\frac{\partial \mathbf{F}(\mathbf{x}(b)|b)}{\partial \mathbf{x}}\right]^{-1} \frac{\partial \mathbf{F}(\mathbf{x}(b)|b)}{\partial b} = \begin{cases} e^{x_2}/(2x_2 + 2x_1 e^{x_2}) \\ 1/(2x_2 + 2x_1 e^{x_2}) \end{cases}.$$

The previous argument was designed to hold for arbitrary systems of equations. Optimization problems however, have a unique feature though, namely that the condition $|D_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}|\bar{\mathbf{b}})| \neq 0$ always holds. This is the case since in an optimization problem such matrix simply coincides with the Hessian of the Lagrangian (that is, the bordered Hessian).

This means that indeed we can take the maximum value function, or set of equilibrium conditions, totally differentiate them and find how the endogenous variables change with the exogenous ones in the neighborhood of the solution.

An Example: Consider a simple optimization with two variables and one equality constraint

$$\max_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq b.$$

First order conditions require

$$\begin{aligned}
F^1(\mu, \mathbf{x}|b) &= g(\mathbf{x}) - b = 0, \\
F^2(\mu, \mathbf{x}|b) &= f_1(\mathbf{x}) - \mu g_1(\mathbf{x}) = 0, \\
F^3(\mu, \mathbf{x}|b) &= f_2(\mathbf{x}) - \mu g_2(\mathbf{x}) = 0.
\end{aligned}$$

We need to verify that the Jacobian of the system is different from zero,

$$|D_{\mu,\mathbf{x}}\mathbf{F}(\mu, \mathbf{x}|b)| = \begin{vmatrix} \frac{\partial F^1}{\partial \mu} & \frac{\partial F^1}{\partial x_1} & \frac{\partial F^1}{\partial x_2} \\ \frac{\partial F^2}{\partial \mu} & \frac{\partial F^2}{\partial x_1} & \frac{\partial F^2}{\partial x_2} \\ \frac{\partial F^3}{\partial \mu} & \frac{\partial F^3}{\partial x_1} & \frac{\partial F^3}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 0 & -g_1 & -g_2 \\ -g_1 & f_{11} - \mu g_{11} & f_{21} - \mu g_{21} \\ -g_2 & f_{12} - \mu g_{12} & f_{22} - \mu g_{22} \end{vmatrix} \neq 0.$$

But the determinant of such a matrix is that of the bordered Hessian of the problem.

Thus, whenever second order conditions are satisfied:
(1) the determinant of the bordered Hessian must be different zero (negative) and;
(2) we can totally differentiate the equations to find how the solution changes with b,

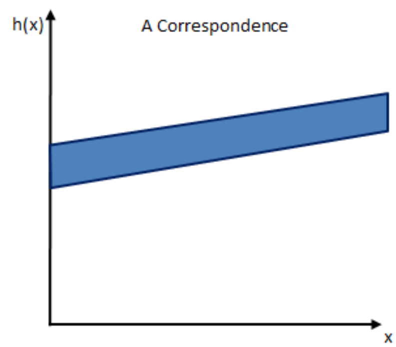
$$\begin{aligned}
g_1 \frac{\partial x_1}{\partial b} + g_2 \frac{\partial x_2}{\partial b} - 1 &= 0, \\
(f_{11} - \mu g_{11}) \frac{\partial x_1}{\partial b} + (f_{12} - \mu g_{12}) \frac{\partial x_2}{\partial b} - g_1 \frac{\partial \mu}{\partial b} &= 0, \\
(f_{21} - \mu g_{21}) \frac{\partial x_1}{\partial b} + (f_{22} - \mu g_{22}) \frac{\partial x_2}{\partial b} - g_2 \frac{\partial \mu}{\partial b} &= 0.
\end{aligned}$$

At the equilibrium solution, one can then solve for $\frac{\partial x_1}{\partial b}$, $\frac{\partial x_2}{\partial b}$, and $\frac{\partial \mu}{\partial b}$.

As equilibria in economics are generally defined as fixed points of some system of equations, proving equilibrium existence essentially relies on classical fixed point theorems. This true in game theory and in general equilibrium theory. We state these results more generally for correspondences (and not just for functions) since the solutions of the optimization problems faced by players may not be unique.

Definition: A *correspondence* from set X to set Y, denoted $h : X \rightrightarrows Y$, is a function from X to $P(Y) = \{A | A \subseteq Y\}$ which is the set of all subsets of Y.

Remark: A correspondence h is set-valued function, as $h(x) \subseteq Y$ for every $x \in X$.



Examples: For instance, the following set-valued functions are correspondences:

- $h(x) = (2, \infty)$;
- $h(x) = [x/2 + 10, x/2 + 11]$;
- $h(x) = \begin{cases} 0 & \text{if } \; x \in [0, 1) \\ [0, 1] & \text{if } \; x = 1 \end{cases}$.

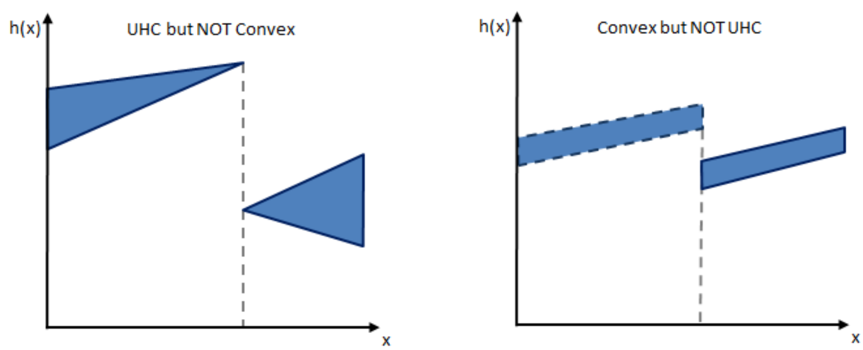
Definition: A correspondence $h : X \rightrightarrows Y$ is:

- *non-empty* if $h(x) \neq \emptyset$ for all $x \in X$;
- *convex-valued* if $y \in h(x)$ and $y' \in h(x)$ imply that

$$py + (1-p)y' \in h(x) \;\; \text{for any} \;\; p \in [0,1];$$

- *upper-hemicontinuous* if for any a sequence $\{y^k, x^k\}_{k=1}^{\infty}$ such that $y^k \in h(x^k)$ for all k,

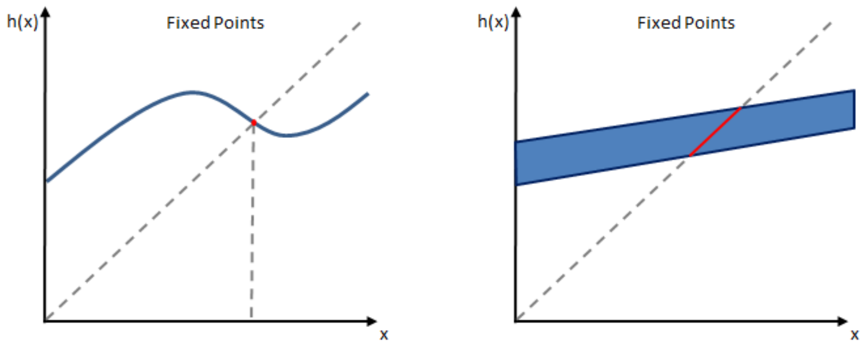
$$x^k \to x \;\; \text{and} \;\; y^k \to y \;\; \text{imply that} \;\; y \in h(x).$$



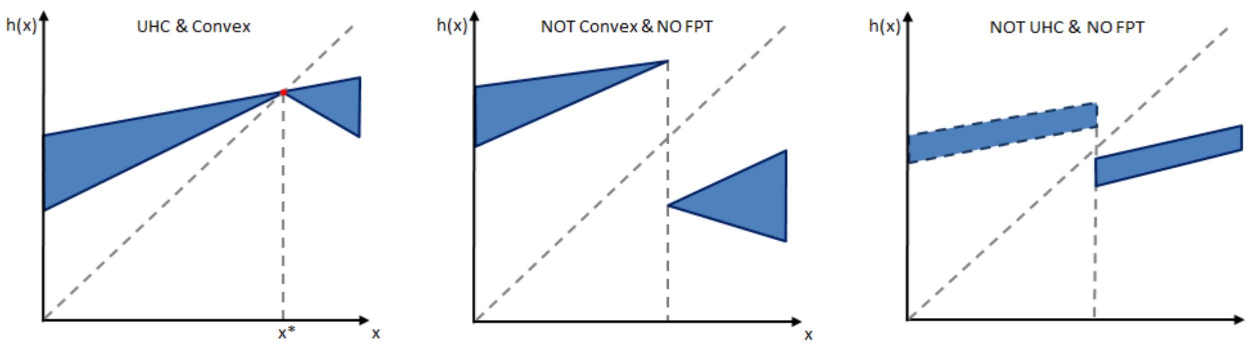
Remark: A correspondence h is upper-hemicontinuous if and only if its graph, $Gr(h) = \{(y,x) \in Y \times X \mid y \in h(x)\}$, is a closed set.

Remark: Any function is a convex-valued correspondence. Any continuous function is a upper-hemicontinuous correspondence.

Definition: A *fixed point* of the correspondence is a point $x \in X$ such that $x \in h(x)$.



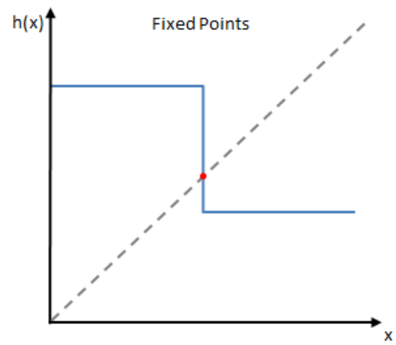
Kakutani Fixed Point Theorem: Let X be a non-empty, compact and convex subset of $\mathbb{R}^n$. Let $h : X \rightrightarrows X$ be a non-empty, convex-valued, upper-hemicontinuous correspondence. Then h has a fixed point.



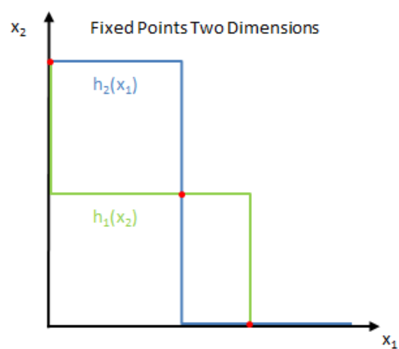
Proof Intuition One-Dimension: Let $X = [0, 1]$ which is a non-empty, compact and convex. If $0 \in h(0)$ or if $1 \in h(1)$, we have a fixed point. If not, $y > 0$ for all $y \in h(0)$ and $y < 1$ for all $y \in h(1)$. If so, take the smallest value x such that $y \leq x$ for some $y \in h(x)$. By UHC, there exists $y' \in h(x)$ such that $y' \geq x$. To show the latter, consider any sequence $\{y^k, x^k\}_{k=1}^{\infty}$ such that $y^k \in h(x^k)$ and $x^k < x$ for all k. If so, as $y^k > x^k$ for all k, we have that

$$y' \equiv \lim_{k\to\infty} y^k \geq \lim_{k\to\infty} x^k \geq x,$$

and by UHC we further have that $y' \in h(x)$. But if so, as h is convex-valued, we must have that $x \in h(x)$.



Two-Dimensions Plot: Consider a two-dimensional correspondence $h(x) \subset \mathbb{R}^2$. Suppose that $h_i(x) = h_i(x_j) \subset \mathbb{R}$. If so, we can exploit the following graph in order to plot the fixed points.

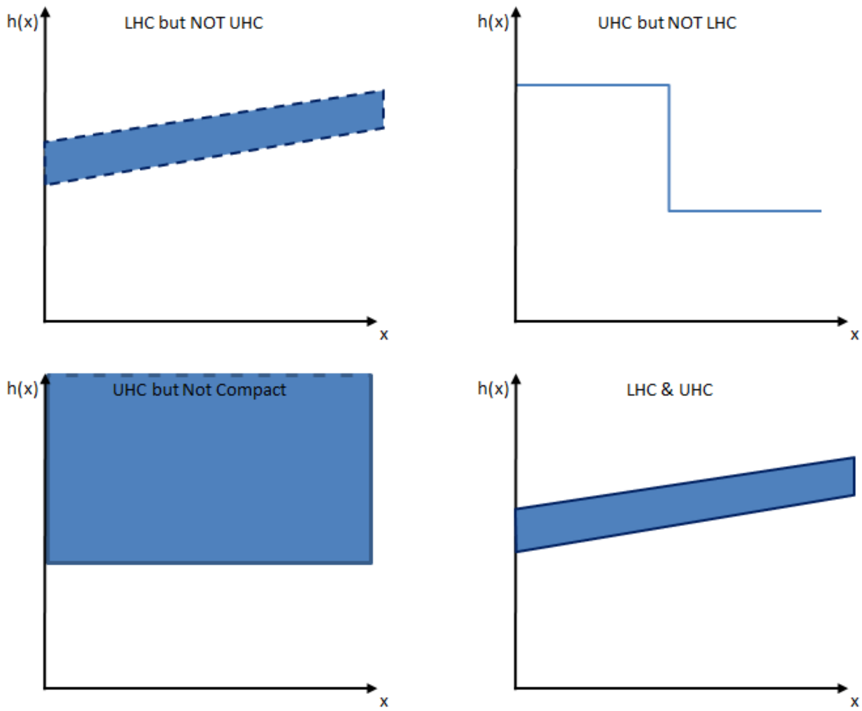


Definition: A correspondence $h : X \rightrightarrows Y$ is:

- *lower-hemicontinuous* if for any $(y, x) \in Gr(h)$ and for any sequence $\{x^k\}_{k=1}^{\infty}$ such that $x^k \to x$, there exists a sequence $\{y^k\}_{k=1}^{\infty}$ such that

$$y^k \in h(x^k) \text{ for all } k \text{ and } y^k \to y;$$

- *continuous* if it is both lower-hemicontinuous and upper-hemicontinuous;
- *compact-valued* if $h(x)$ is compact for all $x \in X$.



Theorem of the Maximum: Let X be a compact subset of $\mathbb{R}^n$, and let B be a subset of $\mathbb{R}^m$. Let $f : X \times B \to \mathbb{R}$ be a continuous function, and let $g : B \rightrightarrows X$ be a continuous correspondence. Let $x^* : B \rightrightarrows X$ be the correspondence defined by

$$x^*(\mathbf{b}) = \arg\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}|\mathbf{b}) \text{ subject to } \mathbf{x} \in g(\mathbf{b}).$$

If so the following must hold:

(1) x^* is *upper-hemicontinuous*;

(2) if g is *non-empty* and *compact-valued*, so is x^*.

Remark: The theorem plays a key role in invoking FPT as it delivers UHC and non-empty compact values *not convex values.*

Remark: If one further invokes that:

- $f(\mathbf{x}|\mathbf{b})$ is quasi-concave in $\mathbf{x}$ for all $\mathbf{b}$;
- $g(\mathbf{b})$ convex-valued for all $\mathbf{b}$;

then x^* is *convex-valued.*

A player starts life with wealth $a_0 > 0$. Each period $t = 0, \ldots, \infty$, he decides how much to consume $c_t \geq 0$. Consumption is chosen to maximize your lifetime utility,

$$\sum_{t=0}^{\infty} \beta^t u(c_t),$$

where $\beta \in [0, 1)$ is the discount factor. Further assume that $\lim_{c \to 0} u'(c) = +\infty$, so that $c_t > 0$ each period. Wealth at the beginning of period t is denoted by a_t. Wealth a_t is invested at a constant interest rate r. Wealth evolves according to the law of motion

$$a_{t+1} = (1 + r)a_t - c_t.$$

Given $a_0 > 0$, the problem is solved by $\{c_t\}_{t=0}^{\infty}$ and $\{a_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t u(c_t).$$

subject to the constraints that for all $t \geq 0$,

$$\begin{aligned} a_{t+1} &= (1 + r)a_t - c_t, \\ a_{t+1} &\geq 0. \end{aligned}$$

To find the solution as a sequence $\{c_t, a_{t+1}\}_{t=0}^{\infty}$ write the Lagrangian:

$$\mathcal{L} = \sum_{t=0}^{\infty} [\beta^t u(c_t) - \mu_t(a_{t+1} - (1 + r)a_t + c_t) - \lambda_t a_{t+1}].$$

where $\{\mu_t, \lambda_t\}_{t=0}^{\infty}$ are sequences of Lagrange multipliers associated to the double sequence of constraints. The first-order conditions of the problem require for all $t \geq 0$:

$$\begin{aligned} \partial\mathcal{L}/\partial c_t &= \beta^t u'(c_t) - \mu_t = 0, \\ \partial\mathcal{L}/\partial a_{t+1} &= (1 + r)\mu_{t+1} - \mu_t - \lambda_t = 0. \end{aligned}$$

Complementary slackness requires instead for all $t \geq 0$:

$$\lambda_t a_{t+1} = 0 \quad \text{and} \quad \mu_t(a_{t+1} - (1 + r)a_t + c_t) = 0.$$

Necessary conditions are also sufficient if the problem is well behaved. Wealth a_t never falls to zero. If $a_T = 0$ for some T, then $c_t = 0$ for all $t \geq T$ which cannot be as $u'(0) = \infty$. Therefore, we have that for all $t \geq 0$:

$$a_{t+1} > 0; \;\; \lambda_t = 0; \;\; (1 + r)\mu_{t+1} = \mu_t.$$

Consequently, for all $t \geq 0$, consumption c_t satisfies:

$$u'(c_t) = (1 + r)\beta u'(c_{t+1}),$$

and, if so, wealth a_t satisfies for all $t \geq 0$:

$$a_{t+1} = (1 + r)a_t - c_t.$$

London School of Economics
Department of Economics

Dr Francesco Nava
Office: 32L.3.20

EC400 – Math for Microeconomics
Problem Set 1: Tools for Optimization

1. Show that the general quadratic form
$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$$
can be written as
$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix};$$
and find its unique symmetric representation.

2. List all the principal minors of a general (3×3) matrix and denote which are the three leading principal minors.

3. Determine the definiteness of the following symmetric matrices:
$$\text{(a)} \begin{pmatrix} 0 & 0 \\ 0 & c \end{pmatrix}; \quad \text{(b)} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}; \quad \text{(c)} \begin{pmatrix} -3 & 4 \\ 4 & -6 \end{pmatrix}; \quad \text{(d)} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{pmatrix}.$$

4. [Harder] Consider the following quadratic form
$$Q(\mathbf{x}) = ax_1^2 + bx_2^2 + 2abx_1x_2.$$
For what values of the parameter values, a and b, is the quadratic form $Q(\mathbf{x})$ indefinite? Plot your answer in $\mathbb{R}^2$.

5. Approximate e^x at $x = 0$ with a Taylor polynomial of order three and four. Then compute the values of these approximation at $h = 0.2$ and at $h = 1$ and compare with the actual values.

6. For each of the following functions on $\mathbb{R}$, determine whether they are quasiconcave, quasiconvex, both, or neither:
$$\text{(a) } e^x; \quad \text{(b) } \ln x; \quad \text{(c) } x^3 - x$$

7. [Harder] Let f be a function defined on a convex set U in $\mathbb{R}^n$. In lecture, we have shown that f is a quasiconcave function on U if and only if for all $\mathbf{x}, \mathbf{y} \in U$ and $t \in [0, 1]$
$$f(t\mathbf{x} + (1-t)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y})\}.$$
State the corresponding theorem for quasiconvex functions and prove it.

8. Prove that a weakly increasing transformation ($g : \mathbb{R} \to \mathbb{R}$ such that $g' \geq 0$) of a quasi-concave function is quasi-concave.

9. A commonly used production or utility function is $f(x, y) = xy$. Check whether it is concave or convex using its Hessian. Then, check whether it is quasiconcave.

10. [Harder] Show that any continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$, satisfying

$$x \frac{\partial f(x)}{\partial x} \leq 0,$$

must be quasi-concave.

London School of Economics
Department of Economics

Dr Francesco Nava
Office: 32L.3.20

EC400 – Math for Microeconomics
Problem Set 2: Optimization

1. Find critical points, local maxima and local minima for each of the following functions:

   (a) $x^4 + x^2 - 6xy + 3y^2$
   (b) $x^2 - 6xy + 2y^2 + 10x + 2y - 5$
   (c) $xy^2 + x^3y - xy$

   Which of the critical points are also global maxima or global minima?

2. Let $S \subset \mathbb{R}^n$ be an open set and $f : S \to \mathbb{R}$ be a twice continuously differentiable function. Suppose that $Df(x^*) = 0$. State the weakest sufficient conditions that the Hessian must satisfy at the critical point x^* for:

   (i) x^* to be a local max;

   (ii) x^* to be a strict local min.

3. Check whether $f(x, y) = x^4 + x^2y^2 + y^4 - 3x - 8y$ is concave or convex by using the Hessian.

4. [Harder] Solve the following problem

$$\max_{x,y} \left[\min\{x, y\} - x^2 - y^2\right].$$

5. Find the optimal solution for the following program

$$\max_{x,y} x \text{ subject to } x^3 + y^2 = 0.$$

   Is the Lagrange approach appropriate?

6. Solve the following problem

$$\max_{x_1,x_2} x_1^2 x_2 \text{ subject to } 2x_1^2 + x_2^2 = 3.$$

7. [Harder] Solve the following problem when $a \in \left[\frac{1}{2}, \frac{3}{2}\right]$

$$\max_{x,y\geq 0} x^2 + y^2 \text{ subject to } ax + y = 1.$$

8. [Harder] Let X be a convex subset of $\mathbb{R}^n$, $f : X \to \mathbb{R}$ a concave function, $g : X \to \mathbb{R}^m$ a convex function, a is a vector in $\mathbb{R}^m$. Consider the following problem

$$\max_{x\in X} f(x) \text{ subject to } g(x) \leq a.$$

   What is the Largrangian for this problem? Prove that the Largangian is a concave function of the choice variable x on X.

9. Consider the problem of maximizing xyz subject to $x+y+z \leq 1$, $x \geq 0$, $y \geq 0$, and $z \geq 0$. Obviously, the three latter constraints do not bind, and we can concentrate only on the first constraint, $x+y+z \leq 1$. Find the solution and the Lagrange multiplier, and show how the optimal value would change if instead the constraint was changed to $x+y+z \leq 9/10$.

10. [Harder] Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying:

$$f(x) = \begin{cases} u(x) & \text{if} \quad g(x) \leq 0 \\ v(x) & \text{if} \quad g(x) \geq 0 \end{cases} .$$

Further, suppose that: (i) $u(x) = v(x)$ if $g(x) = 0$; (ii) u and v are differentiable, strictly concave, and posses maximizer in $\mathbb{R}^n$; and (iii) $g(x)$ is differentiable and strictly convex. Carefully explain how you would solve the problem of maximizing f by choosing $x \in \mathbb{R}^n$.

London School of Economics
Department of Economics

Dr Francesco Nava
Office: 32L.3.20

EC400 – Math for Microeconomics
Problem Set 3: Comparative Statics and Correspondences

1. [Harder] For $a > 0$, consider the problem:

$$\begin{array}{rl} \max_{x,y} & ax + y \\ \text{s.t.} & x^2 + (x-y)^2 \leq 1 \\ & x \geq a,\ y \geq 0. \end{array}$$

   Using the Kuhn-Tucker approach, write down the necessary first order conditions that must be satisfied by the solution of the constrained optimization problem. Are solutions to these conditions also maximizers of the Lagrangian? Solve the constrained optimization problem in terms of a. Find the slope of value function with respect to a, possibly relying on the Envelope Theorem.

2. Let a, x and y be non-negative. Consider the problem of maximizing xy subject to $x^2 + ay^2 \leq 1$. What happens to the optimal value when a increases marginally?

3. [Harder] Assume that the utility function of a consumer satisfies

$$u(x, y) = x + y^{1/2}.$$

   The consumer has a positive income $I > 0$ and faces positive prices $p_x > 0$ and $p_y > 0$. The consumer cannot buy negative amounts of any of the goods.

   (a) Use Kuhn-Tucker to solve the consumer's problem.

   (b) Show how the utility at the maximum depends on I.

   (c) For the case of an interior solution, find how the endogenous variables change when I and p_x change. That is, compute:

$$\frac{\partial x}{\partial I}, \frac{\partial y}{\partial I}, \frac{\partial \mu_0}{\partial I}; \frac{\partial x}{\partial p_x}, \frac{\partial y}{\partial p_x}, \frac{\partial \mu_0}{\partial p_x}.$$

4. [Harder] Consider the system of equations:

$$\mathbf{F}(\mathbf{x}|b) = \left\{ \begin{array}{c} x_1 - e^{x_2} \\ e^{x_1} + x_2 - b^2 \end{array} \right. = \mathbf{0}.$$

   Argue whether you can invoke the implicit function theorem. Then, exploit the theorem where possible to find the slopes of the implicit functions $\mathbf{f}(b)$.

5. For each of the following correspondences $h : \mathbb{R} \rightrightarrows \mathbb{R}$ show whether they are convex-valued, upper-hemicontinuous or lower-hemicontinuous:

   (a) $h(x) = [5x, 10x)$ for $x \in [0, 1]$;

   (b) $h(x) = \{5x, 10x\}$ for $x \in [0, 1]$;

   (c) $h(x) = \begin{cases} 1/2 & \text{for} \quad x \in [0,1) \\ [0,1] & \text{for} \quad x = 1 \end{cases}$.

6. Consider the correspondence $h : \mathbb{R} \rightrightarrows \mathbb{R}$ where

$$h(x) = \begin{cases} [0, 1/2] & \text{if} \quad x > 1/3 \\ [1/2, 1] & \text{if} \quad x < 1/3 \end{cases}.$$

   Argue whether the Kakutani's Fixed Point Theorem applies and find all the fixed points.

7. Consider the correspondence $h : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$, where $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}))$,

$$h_1(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad x_2 > 1/3 \\ [0,1] & \text{if} \quad x_2 = 1/3 \\ 0 & \text{if} \quad x_2 < 1/3 \end{cases} \quad \text{and} \quad h_2(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad x_1 < 1/2 \\ [0,1] & \text{if} \quad x_1 = 1/2 \\ 0 & \text{if} \quad x_1 > 1/2 \end{cases}.$$

   Show how you can expolit Kakutani's Fixed Point Theorem despite the domain being open, and find all the fixed points.

8. [Harder] For all values of $b \geq 0$, solve the problem

$$\mathbf{x}^*(b) = \arg\max_{\mathbf{x} \in [0,1]^2} bx_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 = 1.$$

   Show that $\mathbf{x}^*(b)$ is non-empty, compact-valued, convex-valued, and upper-hemicontinuous. Would this have followed also from the theorem of the maximum? Are the assumptions of the theorem satisfied?