Mathematical review

1. Sets

A **set** is any collection of elements.

Examples:

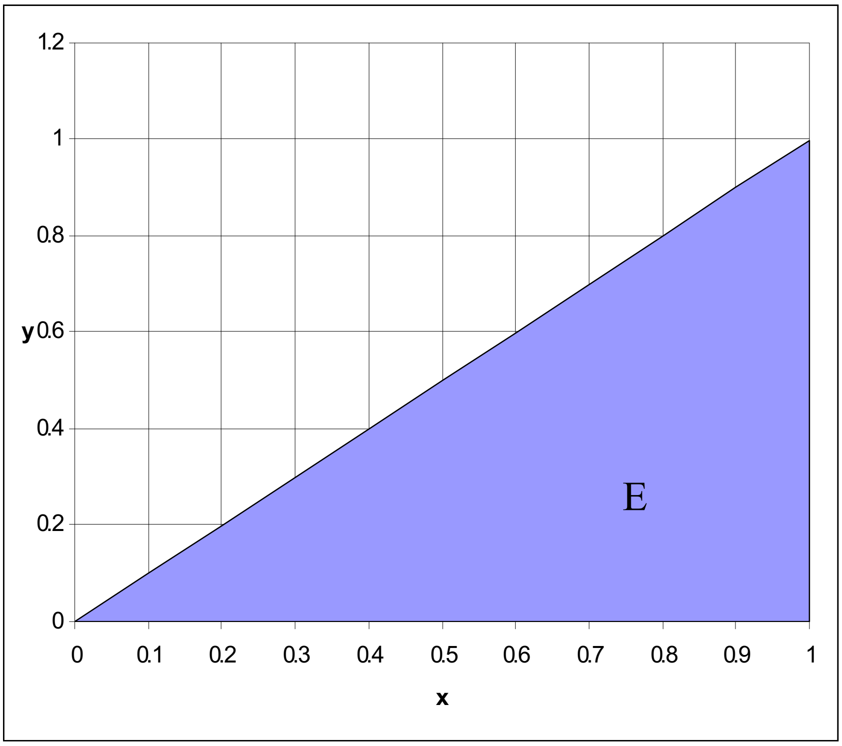
a. $A = \{0,2,4,6,8,10\}$ - the set of even numbers between zero and 10.

b. $B = \{red, white, bule\}$ - the set of colors on the national flag.

c. $C = \{U\ of\ M\ students \mid feamale, GPA \geq 3.2\}$ - the set of U of M students that satisfy the conditions listed after the vertical bar.

d. $D = \{(x, y) \in R^2 \mid x = y\}$ - the set of vectors in the two dimensional Euclidean space, such that the x-coordinate is equal to the y-coordinate.

e. $E = \{(x, y) \in R^2 \mid 0 \leq x \leq 1, 0 \leq y \leq 1, x \geq y\}$ - the set of vectors in the two dimensional Euclidean space such that both coordinates are between 0 and 1 and the x-coordinate is greater or equal to the y-coordinate. It is useful to illustrate this set graphically.



The set E is colored blue.

Cartesian product of A and B is sets the set of all ordered pairs such that the first element belongs to A and the second belongs to B. We denote the Cartesian product by $A \times B$.

Example: $A = \{1,2,3\}$, $B = \{7,8\}$, then $A \times B = \{(1,7),(1,8),(2,7),(2,8),(3,7),(3,8)\}$.

Convex sets: A set B is convex if $\forall x, y \in B \quad \alpha x + (1-\alpha)y \in B \quad \forall \alpha \in [0,1]$. In words, a linear combination of any two elements in the set also belongs to the set.

2. Functions of one variable.

A **function** $f : A \to B$ consists of the *domain* set (A) the *range* set (B) and a rule that assigns to every element in the domain, a <u>unique</u>[1] element in the range. We can say that the function *f* maps from A into B.

Examples:

a. Let $A = \{x \in R \mid 0 \le x \le 100\}$. Let $B = \{A, A-, B+, B, B-, C+, C, C-, D+, D, F\}$. The grading scale at the end of the syllabus is a rule that assigns a unique element in B (a letter grade) to every numerical grade between 0 and 100. We can give this grading function a name - G, and write $G : A \to B$, where A is the domain and B is the range.

b. Let $f : R \to R$ be a function from the set of real numbers into the set of real numbers, such that $f(x) = x^2$. The domain of this function is R, the range is also R and the function assigns to every number its square.

2.1. Graphs of Functions.

The **graph** of a function $f : A \to B$ is a set that all ordered pairs of the form $(x, f(x))$ such that $x \in A$ and $f(x) \in B$. Formally $Gr(f) = \{(x, y) \in A \times B \mid y = f(x)\}$.

Example: let $f : R \to R$, and $f(x) = x^2$. Then $Gr(f) = \{(x, y) \in R \times R \mid y = x^2\}$ is the graph of *f*.

2.2. Linear functions.

The general form of a linear function is $f(x) = a + bx$. Here *a* is the **intercept** with the vertical axis and b is the **slope** of the function. It is important to be able to plot linear functions given their equation. The following formula gives the equation of a linear function based on two points in the plane. Let $(x_1, y_1), (x_2, y_2)$ be two points in R^2. The equation for the line through those points is given by

$$y = \left(\frac{y_1 - y_2}{x_1 - x_2}\right)(x - x_1) + y_1$$

The quotient on the right hand side is the slope pf the line. Hence, we can use this formula to find a linear function when given only one point and the slope. Let (x_1, y_1) be a point in R^2 and given that the slope of a line that passes through this point is *b*, the equation of the line is given by

$$y = b(x - x_1) + y_1$$

2.3. Inverse functions.

Let *f* be a function. Then *f* **has an inverse** if there is a function *g* such that the domain of *g* is the range of *f* and such that

$$f(x) = y \quad \textit{if and only if} \quad g(y) = x$$

for all x in the domain of *f* and all *y* in the range of *f*. In this case, *g* is the **inverse** of *f*, and is designated by f^{-1}.

[1] A mapping that assigns possibly more than one value to every element in the domain is called *correspondence*.

Thus

$$f(x) = y \quad \textit{if and only if} \quad f^{-1}(y) = x$$

for all x in the domain of f and all y in the range of f.
Example: let $f(x) = x^3$. Then $f^{-1}(y) = x^{1/3}$.

2.4. Implicit and explicit functions.

Some times when a function is defined, it is clear which variable is the function of the other, for example: $y: R \to R$, $y(x) = 2x$. Here it is clear that y is a function of x and we call this form **explicit form** or **explicit function.** On the other hand, in some cases it is not clear which variable is the function of the other variable. For example, when we see the equation $y - 2x = 0$, it is not clear that y is a function of x (in which case we can write $y(x) = 2x$) or x is a function of y (in which case we write $x(y) = 0.5y$). We call the form $y - 2x = 0$ **implicit form** or **implicit function.** In general, an implicit function of two variables is of the form $g(x, y) = c$, where c is a constant.

2.5. Derivatives of functions.

The **derivative** of a function at the point x is defined to be $\lim_{h \to 0} \frac{f(x+h) - h}{h}$. If such limit exists, we say that f is **differentiable** at the point x and denote this limit by $\frac{df(x)}{dx}$ or $f'(x)$. If f is differentiable at all point in the domain, we say that f is **differentiable function.** The derivative of a function at x is equal to the slope of the function at x. If we let x vary over the domain, then the derivative itself is a function of x and we call it the **derivative function.**

Important rules of differentiation.

For constants, α : $\frac{d}{dx}(\alpha) = 0.$

For sums: $\frac{d}{dx}[f(x) \pm g(x)] = f'(x) \pm g'(x).$

Power rule: $\frac{d}{dx}(\alpha x^n) = n\alpha x^{n-1}$.

Product rule: $\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$.

Quotient rule: $\frac{d}{d(x)}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$.

Chain rule: $\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$.

Logarithmic function: $\frac{d}{dx}\ln(x) = \frac{1}{x}$.

Exponential function: $\frac{d}{dx}e^x = e^x$

2.6. Second derivatives.

The **second derivative** of a function is the derivative of the *derivative function*. Suppose $y = f(x)$ and the derivative function is $f'(x)$. Then the second derivative of $f(x)$ with respect to *x* is $\frac{d^2 f(x)}{dx^2}$ or $f''(x)$.

Example: $f(x) = x^3$, the (first) derivative function of *f* is $f'(x) = 3x^2$ and the second derivative is $f''(x) = 6x$.

3. Functions of several variables.

Our definition of a function in section 2 was general and we did not restrict the domain to be one-dimensional. Now we focus attention on functions of several variables.

Examples:

a. Let $M : R^3 \times G \to R$, $M = f(w,n,r,g)$ be the function that describes the size of a muscle as a function of (w) workout, (n) nutrition, (r) rest time, and (g) genetic characteristics of the athlete. This is a function of four variables where the first three are assumed to be expressed in terms of real variables and the last one belongs to a set of all possible genetic characteristics for human.

b. Let $\pi : R^2 \to R$, $\pi = f(p,a)$ be the profit function of a firm that depends on the price of the firm's product and the amount of advertising.

3.1. Partial derivatives

Let $y = f(x_1,...,x_n)$. The partial derivative of f with respect to x_i is defined as

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1,...,x_i + h,...,x_n) - f(x_1,...,x_n)}{h}$$

Other notations: $\partial y / \partial x_i$ or $f_i(\mathbf{x})$ or $f_{x_i}(\mathbf{x})$[2].

Examples

a. Let $f(x_1, x_2) = x_1^2 + 3x_1x_2 - x_2^2$. Since this is a function of two variables, the two partial derivatives are: $\frac{\partial f(x_1, x_2)}{\partial x_1} = 2x_1 + 3x_2$ and $\frac{\partial f(x_1, x_2)}{\partial x_2} = 3x_1 - 2x_2$.

b. In the following example you need to use the chain rule. Suppose that $g(x_1, x_2)$ and $x_1(t), x_2(t)$. That is g is a function of x_1, x_2 and those variables are themselves functions of t. Find the derivative $g'(t)$.

Solution: $\frac{dg[x_1(t), x_2(t)]}{dt} = \frac{\partial g[x_1(t), x_2(t)]}{\partial x_1}\frac{dx_1(t)}{dt} + \frac{\partial g[x_1(t), x_2(t)]}{\partial x_2}\frac{dx_2(t)}{dt}$.

Sometimes you will see the derivative with simpler notations (for lazy people like myself): $\frac{dg(t)}{dt} = g_1[\ ]x_1'(t) + g_2[\ ]x_2'(t)$.

---

[2] The bold faced letters denote vectors. That is, $\mathbf{x} = (x_1,...,x_n)$.

3.2. Differential

Let $f(x_1,...,x_n)$ be a function. The **full differential** of this function is defined as follows:

$$df(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1}dx_1 + ... + \frac{\partial f(\mathbf{x})}{x_n}dx_n$$

The differential gives the change in the value of the function f when we change x_1 by dx_1, x_2 by dx_2,…, and x_n by dx_n. The changes $dx_1,...,dx_n$ are assumed to be small.

Example: consider the following profit function $f(x_1,x_2) = 10x_1 - x_1^2 + 20x_2 - x_2^2$, where x_1 is the firm's output and x_2 is the amount of advertising. Suppose that initially the firm produces 4 units of output and 11 units of advertising. Question: What will be the change in the firm's profit if it increases the output by 0.1 units and reduces advertising by 0.05 units? Solution: One way to answer this question is simply calculate the initial profit $f(4,11) = 123$ and the profit after the change $f(4.1,10.95) = 123.2875$. The change in profit is +0.2875. Another way to calculate the approximate change in profit is using the differential.

$$df(x_1,x_2) = \frac{\partial f(\,)}{\partial x_1}dx_1 + \frac{\partial f(\,)}{\partial x_2}dx_2 = (10-2x_1)\cdot 1 + (20-2x_2)\cdot(-0.5) =$$

$$= (10-2\cdot 4)\cdot 0.1 + (20-2\cdot 11)\cdot(-0.05) = 0.2 + 0.1 = 0.3$$

The answer we get with the differential is approximation, and the smaller the changes in the variables, the closer we get to the true change in the value of the function.

3.3. Concave and Convex functions.

$f : A \to R$ is a **concave function** if for all $\mathbf{x}_1, \mathbf{x}_2 \in A$,

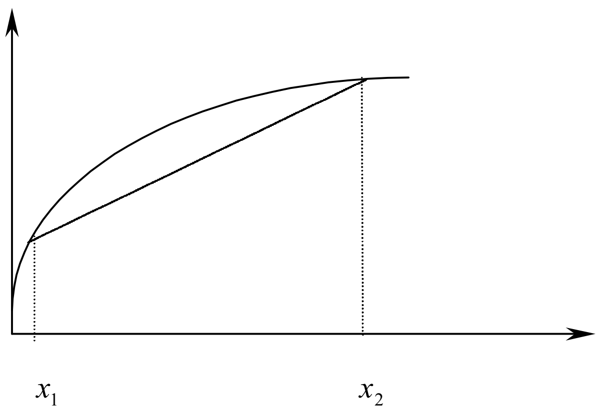
$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \geq \alpha f(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2) \quad \forall\alpha \in [0,1]$$

A function is **strictly concave** if

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) > \alpha f(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2) \quad \forall\alpha \in (0,1)$$

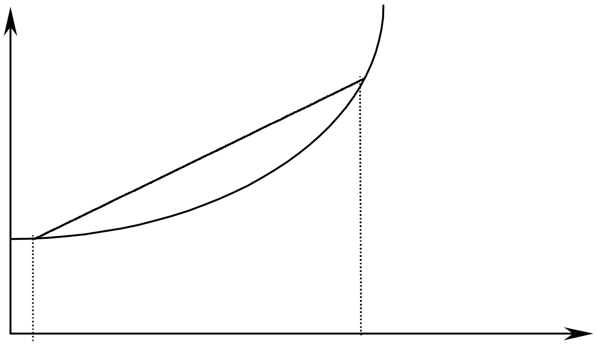
Intuitively, a function is **concave** if for every pair of points on its graph, the cord joining them lies on or below the graph.

<u>Concave function:</u>



For **convex** functions, reverse the inequalities on or above. Intuitively, a function is **convex** if the cord connecting any two points lies on or above the graph of the function.

Convex function:



3.4. Theorem (characterization of concavity with second derivatives).

3.4.1. Let f be twice continuously differentiable function of *one* variable. Then *f* is **concave function** if and only if $f''(x) \le 0$ for all x in the domain of *f*.

3.4.2. Let f be twice continuously differentiable function of *one* variable. Then *f* is **strictly concave function** if and only if $f''(x) < 0$ for all x in the domain of *f*.

For **convex** functions, reverse the inequality. The analogous theorem for multivariate functions is omitted.

3.5. Implicit function theorem (for functions of two variables), sloppy version.

Consider an expression $f(x, y) = c$. This is an implicit form. Now, suppose that we wish to find the derivative $\frac{dy}{dx}$. The implicit function theorem says that if *f* is continuously differentiable, and $\frac{\partial f(x, y)}{\partial y} \neq 0$, then

a. We can solve for $y = y(x)$, i.e., find the explicit form for y.

b. $\frac{dy(x)}{dx} = -\frac{\partial f(x, y) / \partial x}{\partial f(x, y) / \partial y}$.

It should be obvious why the assumption $\frac{\partial f(x, y)}{\partial y} \neq 0$ is necessary for the second part.

The implicit function theorem helps us to find the derivative of *y* with respect to *x* even if we don't have the explicit form for *y* as a function of *x*.

Example: Let $f(x, y) = x^2 + y^2 = r$. Find the derivative $\frac{dy}{dx}$.

Solution: $\frac{dy}{dx} = -\frac{\partial f(x, y) / \partial x}{\partial f(x, y) / \partial y} = -\frac{2x}{2y} = -\frac{x}{y}$.

4. Homogeneous functions

A function $f: R^n \to R$ is **homogeneous of degree *k*** if $f(\lambda \mathbf{x}) = \lambda^k f(\mathbf{x}) \quad \forall \lambda > 0$.

Examples:

a. Let $f(x_1, x_2) = x_1^{0.5} x_2^{0.5}$. This function is homogeneous of degree 1 since $f(\lambda x_1, \lambda x_2) = (\lambda x_1)^{0.5} (\lambda x_2)^{0.5} = \lambda x_1^{0.5} x_2^{0.5} = \lambda f(x_1, x_2)$.

b. Let $f(x_1, x_2) = x_1 x_2$. This function is homogeneous of degree 2 since $f(\lambda x_1, \lambda x_2) = (\lambda x_1)(\lambda x_2) = \lambda^2 x_1 x_2 = \lambda^2 f(x_1, x_2)$

c. Let $f(x_1, x_2) = x_1 / x_2$. This function is homogeneous of degree 0 since $f(\lambda x_1, \lambda x_2) = \lambda x_1 / \lambda x_2 = x_1 / x_2 = \lambda^0 f(x_1, x_2) = f(x_1, x_2)$.

d. Let $f(x_1, x_2) = x_1 + x_2^2$. This function is not homogeneous.

4.1. Euler's theorem for homogeneous functions

Theorem: Let $f: R^n \to R$ be homogeneous of degree *k*. Then

$$f_1(\mathbf{x}) x_1 + f_2(\mathbf{x}) x_2 + \ldots + f_n(\mathbf{x}) = k \lambda^{k-1} f(\mathbf{x})$$

Proof: By definition of homogeneous function of degree k

$$(1) \quad f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda^k f(x_1, x_2, \ldots, x_n) \quad \forall \lambda > 0$$

Take the derivative of both sides with respect to λ:

$$(2) \quad f_1(\lambda \mathbf{x}) x_1 + f_2(\lambda \mathbf{x}) x_2 + \ldots + f_n(\lambda \mathbf{x}) x_n = k \lambda^{k-1} f(x_1, x_2, \ldots, x_n)$$

Since (1) is true for all $\lambda > 1$ it must be true for $\lambda = 1$, which gives us the required result.

5. Optimization.[3]

5.1. Unconstrained optimization with one variable.

A function f has a **local maximum** at the point x_0 if there is some interval that contains x_0 and $f(x_0) \geq f(x)$ for all x in that interval.

A function f has a **global maximum** at the point x_0 if $f(x_0) \geq f(x)$ for all x the domain of *f*.

If a function *f* has maximum at x_0, then x_0 is called a **maximizer** of *f*. We write $x_0 = \arg\max f(x)$, which means that x_0 is an argument that maximizes the function *f*. The value of *f* at x_0 is called the **maximum** of *f* or the **maximum value** of *f*.

5.1.1. Theorem (necessary conditions for maximum)

Let $f(x)$ be twice continuously differentiable function of one variable. If *f* has a **local maximum** at x_0 then

a. $f'(x_0) = 0$ (First Order Necessary Condition, FONC).

b. $f''(x_0) \leq 0$ (Second Order Necessary Condition).

For **local minimum**, reverse the inequality in b.

[3] In this section we restrict attention to differentiable function only.

What do we mean by necessary conditions? We mean that the above conditions <u>must</u> be satisfied at the point of maximum, but they cannot guarantee maximum. If we check those conditions for some function and we find that they are satisfied for some point, we CANNOT conclude that we have a maximum at that point.
Example: let $f(x) = x^3$. At $x = 0$ the above conditions are satisfied, but this function is monotone increasing and has neither maximum nor minimum.

5.1.2. Theorem (sufficient conditions for maximum).

Let $f(x)$ be twice continuously differentiable function of one variable. If $f'(x_0) = 0$ and $f''(x_0) < 0$, then *f* has a **local maximum** at x_0.
For **local minimum,** reverse the last inequality.
The above two conditions ($f'(x_0) = 0$ and $f''(x_0) < 0$) are together sufficient for maximum. This means that if they are satisfied at some point of the domain, then the function has local maximum at that point.
Example: Let $f(x) = 10x - x^2$. Verify that *f* has a local maximum at $x_0 = 5$.
Solution: $f'(x) = 10 - 2x, \quad f''(x) = -2$. $\Rightarrow f'(5) = 0, \quad f''(5) = -2 < 0$.

5.2. Unconstrained optimization with two variables

5.2.1. Theorem (sufficient conditions for maximum)

Let $f(x_1, x_2)$ be twice continuously differentiable. If $f_i(x_1^*, x_2^*) = 0 \quad for \quad i = 1,2$ and $f_{11}(x_1^*, x_2^*) < 0$ and $f_{11} \cdot f_{22} - f_{12} \cdot f_{21} > 0$ [4], then *f* has **local maximum** at (x_1^*, x_2^*).

5.2.2. Theorem (sufficient conditions for minimum)

Let $f(x_1, x_2)$ be twice continuously differentiable. If $f_i(x_1^*, x_2^*) = 0 \quad for \quad i = 1,2$ and $f_{11}(x_1^*, x_2^*) > 0$ and $f_{11} \cdot f_{22} - f_{12} \cdot f_{21} > 0$, then *f* has **local minimum** at (x_1^*, x_2^*).
Example: Find a critical point of the function $f(x_1, x_2) = x_2 - 4x_1^2 + 3x_1x_2 - x_2^2$ and check whether it is maximum or minimum.
Solution: We remember form calculus that the point at which all partial derivatives are equal to zero is a critical point.

$$\frac{\partial f}{\partial x_1} = -8x_1 + 3x_2 = 0$$

$$\frac{\partial f}{\partial x_{21}} = 1 + 3x_1 - 2x_2 = 0$$

Solving this system gives us $x_1 = \frac{3}{7}$ and $x_2 = \frac{8}{7}$ (critical point).
Now find the second partial derivatives:
$f_{11} = -8 < 0$, $f_{12} = 3$, $f_{21} = 3$, $f_{22} = -2$.
$\Rightarrow \quad f_{11} \cdot f_{22} - f_{12} \cdot f_{21} = -8 \cdot (-2) - 3 \cdot 3 = 7 > 0$
Hence, the sufficient conditions for maximum are satisfied.

[4] By Young's theorem, for any twice continuously differentiable function $f_{12} = f_{21}$, therefore I could have written the last inequality as $f_{11} \cdot f_{22} - f_{12}{}^2 > 0$

5.3. Constrained optimization

This section is the most important for this course. In fact, most of the problems in this course and in economics in general, are problems of constrained optimization. It is crucial to understand that constrained optimization usually gives entirely different results than unconstrained optimization. Suppose that in one case you can chose any car in the world and in another case you can choose any car who's price is blow $6000. It is likely that your choice would be different in both cases.

I demonstrate two methods for solving constrained optimization problems, through a particular example. Suppose you have 20 fit of rope and you need to construct a rectangular frame with maximal area. This is a problem of constrained maximization. Solution. First, we write this problem in a precise mathematical language:

$$\begin{cases} \max_{x,y} f(x,y) = x \cdot y \\ s.t. \\ 2x + 2y = 20 \end{cases}$$

f is called the **objective function**. In our case *f* represents the area of the rectangular, with the sides *x* and *y*.

Under the "max" we write the **choice variables**. In our case these are the lengths of the sides. "s.t." means "such that" or "subject to". After "s.t." follows the constraint, which in our case means that the sum of lengths of all sides is equal to 20 fit.
It is very important in this course to be able write the problems in a clear way, such that a mathematician, who doesn't know economics, will understand what is written.

5.3.1. Substitution method

The idea is to substitute the constraint into the objective function. First solve for y in the constraint: $y = 10 - x$, and then plug this in the objective function. The resulting optimization problem is

$$\max_{x} x(10 - x) = 10x - x^2 .$$

Notice that this is maximization problem with one variable.
First Order Necessary Condition: $10 - 2x = 0$
Second Order Necessary Condition: $f''(x) = -2 < 0$
The necessary and sufficient conditions for maximum are satisfied. The first order condition yields $x^* = 5$. Substituting into the constraint we find the optimal value for *y*: $y^* = 5$. Hence, the **maximizer** of the objective function is: $(x^*, y^*) = (5,5)$. This means that we should use the rope to construct a squared frame. The **maximum value** of the objective function is $5 \cdot 5 = 25$ squared fit.

5.3.2. Lagrange method

The following theorem is a special case of Lagrange theorem for maximization of functions of two variables with single equality constraint. Obviously, there are generalizations to arbitrary number of variables and inequality constraints.

Theorem: Let *f* and *g* be continuously differentiable functions of two variables. Suppose that (x_1^*, x_2^*) is the solution of the problem

$$\begin{cases} \max_{x_1,x_2} f(x_1, x_2) \\ s.t. \\ g(x_1, x_2) = c \end{cases}$$

Suppose further that (x_1^*, x_2^*) is not a critical point[5] of g. Then, there is a real number λ^* such that $(x_1^*, x_2^*, \lambda^*)$ is a critical point of the Lagrange function

$$L(x_1, x_2, \lambda) \equiv f(x_1, x_2) - \lambda[g(x_1, x_2) - c]$$

In other words, at $(x_1^*, x_2^*, \lambda^*)$

$$\frac{\partial L}{x_1} = 0, \quad \frac{\partial L}{x_2} = 0, \quad \frac{\partial L}{\lambda} = 0.$$

The theorem allows us to convert constrained optimization problem with two unknowns, into an unconstrained optimization, but with three unknowns: x_1, x_2, λ. Notice that the theorem applies to constrained maximization problems, as well as to constrained minimization problems. The theorem provides the first order necessary conditions for optimum.

The number λ is called **Lagrange multiplier** or **shadow price**.

Now, lets apply the theorem to our problem with the rope. The Lagrange function is

$$L(x, y, \lambda) \equiv x \cdot y - \lambda[2x + 2y - 20]$$

The first order necessary conditions are:

(1) $\frac{\partial L}{x} = y - 2\lambda = 0,$

(2) $\frac{\partial L}{x_2} = x - 2\lambda = 0$

(3) $\frac{\partial L}{\lambda} = 2x + 2y - 20$

Notice that the last condition just repeats the constraint.

From (1) and (2) it follows that $x = y$. Plug this in the constraint to get $x^* = y^* = 5$, which means that in order to maximize the area of the rectangular, we should make it a square. The value of the Lagrange multiplier is 2.5.

Remark: there is an economic meaning to the Lagrange multiplier. λ^* gives the change of the optimal value of the objective function that results from small unit change in the constraint. To understand this, suppose that we have 1 more fit of rope to construct the rectangular. What would be the change in the maximal area? According to the Lagrange multiplier, it should increase by 2.5. Indeed, Our calculations say that we should form a square from 21 fit. The side of the square is 5.25 and its area is 27.5625. The Lagrange

[5] Critical point is a point at which all partial derivatives are equal to zero.

multiplier gave us a pretty good approximation for the answer, and the smaller the change in the constraint, the better the approximation.

Second Order Sufficient Condition:
Form the Hessian matrix for the Lagrange function:

$$H(L) = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix}$$

The Second order sufficient condition for maximum says that the determinant of this Hessian should be positive (negative for minimum).