EC5110: MICROECONOMICS

LECTURE 0: PREREQUISITES & OPTIMIZATION

Evan Piermont

Autumn 2018

Front Matter:

- Time:
  - Lectures: Tuesdays 11:00–13:00
  - Seminar: Tuesdays 10:00–11:00
- Office hours: Tuesdays 15:00 – 17:00
- Book: W. Nicholson and C. Snyder. Microeconomic theory: Basic principles and extensions. Nelson Education, 2011. ISBN 9781111525514
- Syllabus: `https://goo.gl/PZhHVD`

What is Microeconomics:

- The study of individual agents making decisions.
- Agents can be:
  - Single Humans
  - Firms, Companies, Universities, etc.
  - Govenments
- Each agent has desires (goals) and actions she can take.

We will use models to study individual decisions, markets, and resources allocation. A **model** is

- A set of assumptions and predictions.
  - We examine how our assumptions map into predictions.
- Simplification of the world.
  - Try to capture universal properties.
  - Exclude inessential complexity.
- Abstractions.
  - Things we learn in one scenario will carry over to others.

We can test models by

- Verifying the assumptions hold.
- Testing the predictions (i.e., statistical analysis).

We should always be thinking about the value of our assumptions!

What is special about Economic Thinking:

- The study of marginal considerations.
- Efficiency can be characterized by marginal conditions.

Roadmap

We will build a theory of resource allocation by markets.

1. Demand: How do consumers decide what to buy?
2. Supply: How do producers decide what to supply?
3. Markets: How do markets form to allow exchange?

But first, an unfortunate foray into math tools: Optimization!

$\mathbb{R}^n$

Most of the course will take place $\mathbb{R}^n$ the n-dimensional Euclidean space.

- Vectors $\boldsymbol{x} \in \mathbb{R}^n$ are n real numbers:

$$\boldsymbol{x} = (x_1, \ldots x_n)$$

- We have addition, scalar multiplication, and an inner product:
  - $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$.
  - for $a \in \mathbb{R}$ let $a\boldsymbol{x} = (ax_1, \ldots, ax_n)$.
  - $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_n x_i y_i$.
- We have the ordering
  - $\boldsymbol{x} \geq \boldsymbol{y}$ if $x_1 \geq y_1 \ldots x_n \geq y_n$.
  - $\boldsymbol{x} > \boldsymbol{y}$ if $x_1 \geq y_1 \ldots x_n \geq y_n$ with some strict.
  - $\boldsymbol{x} \gg \boldsymbol{y}$ if $x_1 > y_1 \ldots x_n > y_n$.

If B is a collection of vectors, then we say that B is convex if for all $\boldsymbol{x}, \boldsymbol{y} \in B$, $s\boldsymbol{x} + (1-s)\boldsymbol{y} \in B$ for any $s \in [0,1]$.

Functions on $\mathbb{R}^n$

$f\colon \mathbb{R}^n \to \mathbb{R}^m$ is linear if

$$f(s\boldsymbol{x} + t\boldsymbol{y}) = sf(\boldsymbol{x}) + tf(\boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

- Linear functions are generalizations of lines.
- It is easy to see that if f is linear then $f(0) = 0$ (why?)
- If $f\colon \mathbb{R}^n \to \mathbb{R}$ is linear than for any finite collection of m vectors, we have $f(\sum_m t_i\boldsymbol{x}_i) = \sum_m t_i f(\boldsymbol{x}_i)$.

Functions on $\mathbb{R}^n$

$f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if and only if

$$f(s\boldsymbol{x} + (1-s)\boldsymbol{y}) = sf(\boldsymbol{x}) + (1-s)f(\boldsymbol{y})$$

for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

- Clearly all linear functions are affine.
- Affine functions are also generalizations of lines, but they do not necessarily pass through the origin.

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
- $f(x_1, x_2) = x_1 x_2$
- $f(x_1, x_2) = 4x_1 + 3x_2$
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
- $f(x_1, x_2) = \ln(x_1 + x_2)$

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
  - Linear
- $f(x_1, x_2) = x_1 x_2$
- $f(x_1, x_2) = 4x_1 + 3x_2$
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
- $f(x_1, x_2) = \ln(x_1 + x_2)$

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
    - Linear
- $f(x_1, x_2) = x_1 x_2$
    - Neither
- $f(x_1, x_2) = 4x_1 + 3x_2$
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
- $f(x_1, x_2) = \ln(x_1 + x_2)$

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
  - Linear
- $f(x_1, x_2) = x_1 x_2$
  - Neither
- $f(x_1, x_2) = 4x_1 + 3x_2$
  - Linear
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
- $f(x_1, x_2) = \ln(x_1 + x_2)$

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
  - Linear
- $f(x_1, x_2) = x_1 x_2$
  - Neither
- $f(x_1, x_2) = 4x_1 + 3x_2$
  - Linear
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
  - Affine
- $f(x_1, x_2) = \ln(x_1 + x_2)$

Are the following linear, affine (but not linear) or neither:

- $f(x_1, x_2) = x_1$
  - Linear
- $f(x_1, x_2) = x_1 x_2$
  - Neither
- $f(x_1, x_2) = 4x_1 + 3x_2$
  - Linear
- $f(x_1, x_2) = 4x_1 + 3x_2 + 2$
  - Affine
- $f(x_1, x_2) = \ln(x_1 + x_2)$
  - Neither

Functions on $\mathbb{R}^n$

If $f : \mathbb{R}^n \to \mathbb{R}$ is a function, and k is an integer, then f is said to be homogeneous of degree k if

$$f(a\boldsymbol{x}) = a^k f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ and $\alpha > 0$.

- Homogeneity is a generalization of linearity
  - All linear functions are homogenous of degree 1

Are the homogeneous, and if so, of what degree?:

- $f(x_1, x_2) = \max\{x_1, x_2\}$

- $f(x, y, z) = x^5 y^2 z^3$

- $f(x) = ln(x)$ (defined over R_+)

- $f(x, y) = (x^2 + y^2)^{\frac{1}{2}}$

Are the homogeneous, and if so, of what degree?:

- $f(x_1, x_2) = \max\{x_1, x_2\}$
  - H.d.1
- $f(x, y, z) = x^5 y^2 z^3$
- $f(x) = ln(x)$ (defined over R_+)
- $f(x, y) = (x^2 + y^2)^{\frac{1}{2}}$

Are the homogeneous, and if so, of what degree?:

- $f(x_1, x_2) = \max\{x_1, x_2\}$
  - H.d.1
- $f(x, y, z) = x^5 y^2 z^3$
  - H.d.10
- $f(x) = ln(x)$ (defined over R_+)
- $f(x, y) = (x^2 + y^2)^{\frac{1}{2}}$

Are the homogeneous, and if so, of what degree?:

- $f(x_1, x_2) = \max\{x_1, x_2\}$
  - H.d.1
- $f(x, y, z) = x^5 y^2 z^3$
  - H.d.10
- $f(x) = ln(x)$ (defined over R_+)
  - Not homogeneous
- $f(x, y) = (x^2 + y^2)^{\frac{1}{2}}$

Are the homogeneous, and if so, of what degree?:

- $f(x_1, x_2) = \max\{x_1, x_2\}$
  - H.d.1
- $f(x, y, z) = x^5 y^2 z^3$
  - H.d.10
- $f(x) = ln(x)$ (defined over R_+)
  - Not homogeneous
- $f(x, y) = (x^2 + y^2)^{\frac{1}{2}}$
  - H.d.1

Remark.

If f is differentiable and homogenous of degree k, then $\frac{\partial f}{\partial x_i}$ is homogenous of degree $k-1$.

- Differentiate both sides of $f(a\boldsymbol{x}) = a^k f(\boldsymbol{x})$ with respect to x_i.

Functions on $\mathbb{R}^n$

$f : \mathbb{R}^n \to \mathbb{R}^m$ is convex if

$$f(s\boldsymbol{x} + (1-s)\boldsymbol{y}) \leq sf(\boldsymbol{x}) + (1-s)f(\boldsymbol{y})$$

for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and strictly so if the inequality is strict.

$f : \mathbb{R}^n \to \mathbb{R}^m$ is concave if

$$f(s\boldsymbol{x} + (1-s)\boldsymbol{y}) \geq sf(\boldsymbol{x}) + (1-s)f(\boldsymbol{y})$$

for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and strictly so if the inequality is strict.

Functions on $\mathbb{R}^n$

Relatedly, a function is quasi-convex if

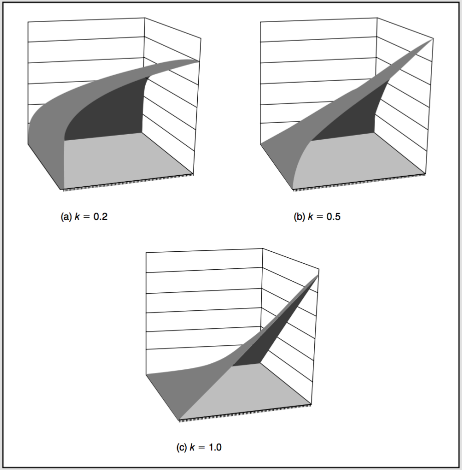
$$f(s\boldsymbol{x} + (1-s)\boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$$

for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and strictly so if the inequality is strict.

Quasi-concave can be defined using pattern matching skills!

FIGURE 2.4 Concave and Quasi-Concave Functions

In all three cases these functions are quasi-concave. For a fixed y, their level curves are convex. But only for $k = 0.2$ is the function strictly concave. The case $k = 1.0$ clearly shows nonconcavity because the function is not below its tangent plane.



Are the following convex, concave, quasi-convex, quasi-concave:

- $f(x_1, x_2) = x_1^2$

- $f(x_1, x_2) = x_1 + x_2$

- $f(x) = ln(x)$ (defined over R_+)

Are the following convex, concave, quasi-convex, quasi-concave:

- $f(x_1, x_2) = x_1^2$
  - Convex, quasi-convex
- $f(x_1, x_2) = x_1 + x_2$
- $f(x) = ln(x)$ (defined over R_+)

Are the following convex, concave, quasi-convex, quasi-concave:

- $f(x_1, x_2) = x_1^2$
  - Convex, quasi-convex
- $f(x_1, x_2) = x_1 + x_2$
  - All (its linear)
- $f(x) = ln(x)$ (defined over R_+)

Are the following convex, concave, quasi-convex, quasi-concave:

- $f(x_1, x_2) = x_1^2$
  - Convex, quasi-convex
- $f(x_1, x_2) = x_1 + x_2$
  - All (its linear)
- $f(x) = ln(x)$ (defined over R_+)
  - Concave, quasi-concave, and quasi-convex.

Remark.

Every convex function is quasi-convex.

By the definition of max

$$\begin{aligned} f(s\boldsymbol{x} + (1-s)\boldsymbol{y}) &\leq sf(\boldsymbol{x}) + (1-s)f(\boldsymbol{y}) \\ &\leq s\max\{f(\boldsymbol{x}), f(\boldsymbol{y})\} + (1-s)\max\{f(\boldsymbol{x}), f(\boldsymbol{y})\} \\ &= \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}. \end{aligned}$$

Remark.

Let f be a quasi-convex function, then $\{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq a\}$ is a convex set for all $a \in \mathbb{R}$.

Optimization.

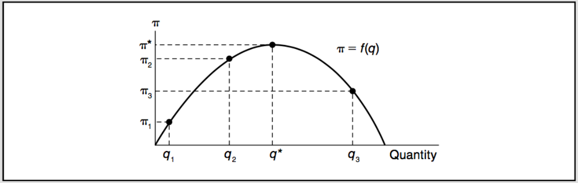
If we have a (well behaved) function, how do we find the optima?

Ex. profit, π depends only on the quantity produced, $q \in \mathbb{R}$ via

$$\pi = f(q)$$

FIGURE 2.1 Hypothetical Relationship between Quantity Produced and Profits

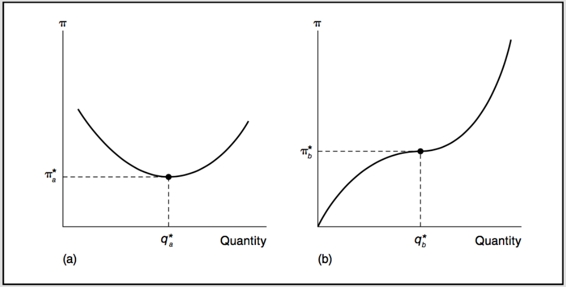
If a manager wishes to produce the level of output that maximizes profits, then q^* should be produced. Notice that at q^*, $d\pi/dq = 0$.



- The slope at the optimum is 0, we need the first derivative to be 0.
- $\frac{\partial f}{\partial q} = 0.$

FIGURE 2.2 Two Profit Functions That Give Misleading Results If the First Derivative Rule Is Applied Uncritically

In (a), the application of the first derivative rule would result in point q_a^* being chosen. This point is in fact a point of minimum profits. Similarly, in (b), output level q_b^* would be recommended by the first derivative rule, but this point is inferior to all outputs greater than q_b^*. This demonstrates graphically that finding a point at which the derivative is equal to 0 is a necessary, but not a sufficient, condition for a function to attain its maximum value.



- $f_q \equiv \frac{\partial f}{\partial q} = 0$ does not guarantee a maximum.

If $f : \mathbb{R} \to \mathbb{R}$ then

- First Order Condition: $\frac{\partial f}{\partial q} = 0$
- Second Order Condition: $\frac{\partial^2 f}{\partial q^2} < 0$

What about functions from $\mathbb{R}^n \to \mathbb{R}$. We can generalize the notion of a derivative and first and second order conditions.

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then we can define the partial derivative of f with respect to x_i as $\frac{\partial f}{\partial x_i}$.

- We treat the other dimensions as constant inputs on which the derivative will depend.
- This is the slope of the function in the i^{th} direction.

Let $f(x_1, x_2) = x_1 x_2^2$ then

- $\frac{\partial f}{\partial x_1} = x_2^2$
- $\frac{\partial f}{\partial x_2} = 2x_1 x_2$

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then the gradient of f is

$$\nabla f = \left(\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_n} \right)$$

is an element of $\mathbb{R}^n$.

The first order condition is easy:

$$\nabla f = (0, \ldots, 0)$$

The second order partial derivative of f with respect to x_i then x_j is as $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

Let $f(x_1, x_2) = x_1 x_2^2$ then

- $\frac{\partial^2 f}{\partial x_1 \partial x_1} = 0$
- $\frac{\partial^2 f}{\partial x_1 \partial x_2} = 2x_2$
- $\frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_2$
- $\frac{\partial^2 f}{\partial x_2 \partial x_2} = 2x_1$

Remark.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then the hessian of f is

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

The second order condition states that the Hessian needs to be negative semi-definite at a maximum.

- Generalization of being negative to matrices.
- Don’t worry about it for this class, but keep in mind in general.

Remark.

If f is concave then the first order condition is also sufficient for maximization!

- Convex :: Minimization

Remark.

If $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ is strictly increasing then

$$\arg\max_{\boldsymbol{x}} f(\boldsymbol{x}) = \arg\max_{\boldsymbol{x}} h(f(\boldsymbol{x}))$$

Parameters

f determines the optimal decision of an agent:

- f depends not only on the choice of the agent, but also on some external parameters which the agent takes as given.
    - For a consumer who has a fixed budget, the parameters might be prices.
    - For a firm deciding how much to produce: the marginal cost of each good.
    - For the government regulating pollution: the environmental cost of emissions.

Example

Consider a firm with a two inputs, k and l. The input is turned into output at a rate of $f(k, l) = \ln(k, l)$.

The firm can sell the output at price p and buy the input at price c_k and c_l.

Profit is given by

$$\pi(k, l, p, c_k, c_l) = p \ln(k, l) - c_k k - c_l l$$

Example

We know that optimal profit must satisfy our first order conditions:

$$p\frac{\partial f}{\partial k} - c_k = \frac{p}{k} - c_k = 0 \tag{π_k}$$

$$p\frac{\partial f}{\partial l} - c_l = \frac{p}{l} - c_l = 0 \tag{π_l}$$

Therefore $k^\star = \frac{p}{c_k}$ and $l^\star = \frac{p}{l_k}$.

Example

Given $k^\star = \frac{p}{c_k}$ and $l^\star = \frac{p}{l_k}$, optimal profit is

$$\pi^\star = p\ln(\frac{p^2}{c_k c_l}) - 2p$$

Comparative Statics

Comparative Statics are the study of how optimal quantities change in response to changes in parameters.

For our example: how does the firms profit change when the cost of labour changes?

Optimal profit: $\pi^\star = p\ln(\frac{p^2}{c_k c_l}) - 2p$.

$\frac{\partial \pi^\star}{\partial c_l} = -\frac{p}{c_l} = -l^\star$

Can we just differentiate π with respect to l?

- $\frac{\partial \pi}{\partial c_l} = -l$
- Ignores that the optimal levels of l and k depend on prices?
- This is okay!
  - Change in quantity of labour has no effect on profit.
  - This is the FOC!

Comparative Statics

Generally: $f\colon \mathbb{R}^{n+k} \to \mathbb{R}$.

- n quantities need to be chosen by the agent
- the problem depends on k constraints

Comparative Statics

Consider the map $x^\star : \mathbb{R}^k \to \mathbb{R}^n$:

$$x^\star : \boldsymbol{a} \mapsto \arg\max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}, \boldsymbol{a}).$$

and the map $f^\star : \mathbb{R}^k \to \mathbb{R}$:

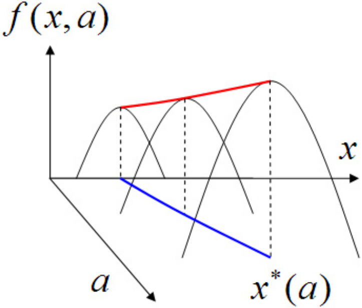
$$f^\star : \boldsymbol{a} \mapsto f(x^\star(\boldsymbol{a}), \boldsymbol{a}).$$

Comparative Statics

Comparative Statics are the study of how $x^\star$ and $f^\star$ responds to changes in $\boldsymbol{a}$.

Theorem. (The Envelope Theorem)

Let $f\colon \mathbb{R}^{n+k} \to \mathbb{R}$. Then so long as both partial derivatives exist, we have $\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{a}}$ (evaluated at the optimum).



$f(x,a)$
x
a
$x^*(a)$

Constrained Optimization

So far, we have dealt with functions which can be optimized over the entire domain; to be concrete, in the above example, the firm can choose *any* allocation of inputs. Many economic situations are constrained

How do we optimize with respect to such constraints?

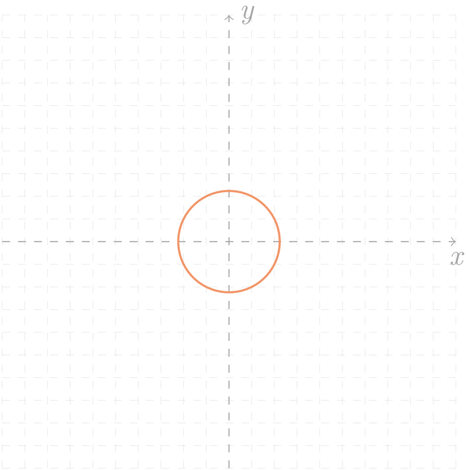
Example

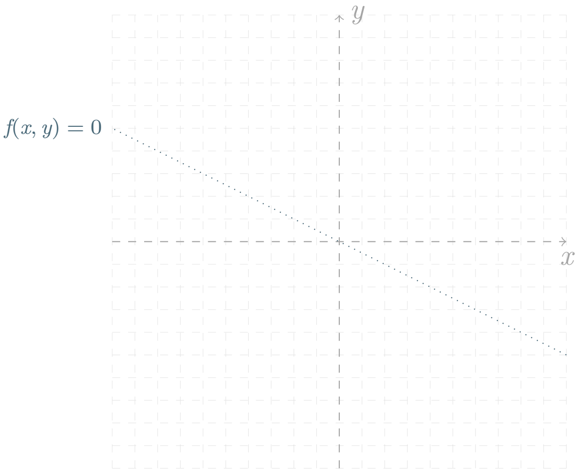
We want to maximize

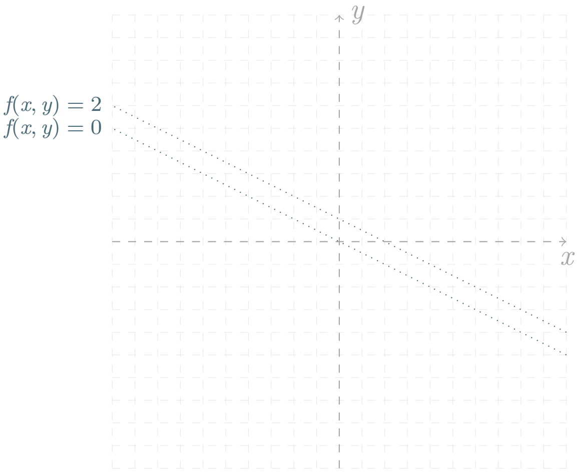
$$f(x, y) = x + 2y$$

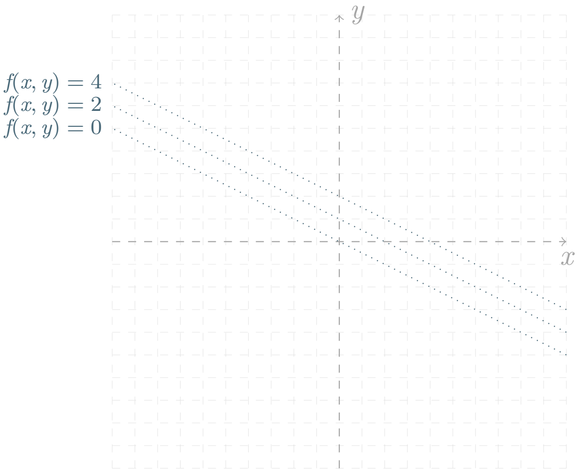
But we must have:

$$x^2 + y^2 = 5$$

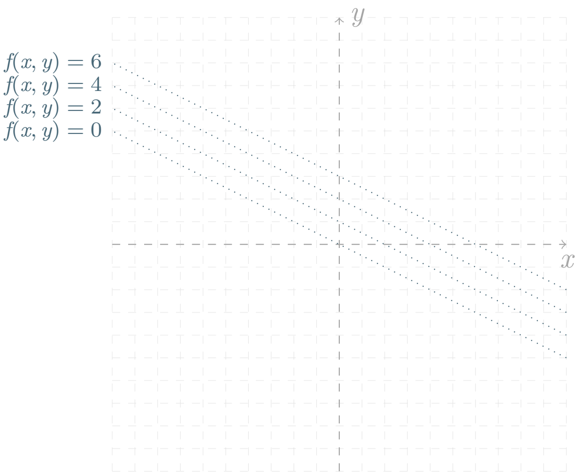

y
x

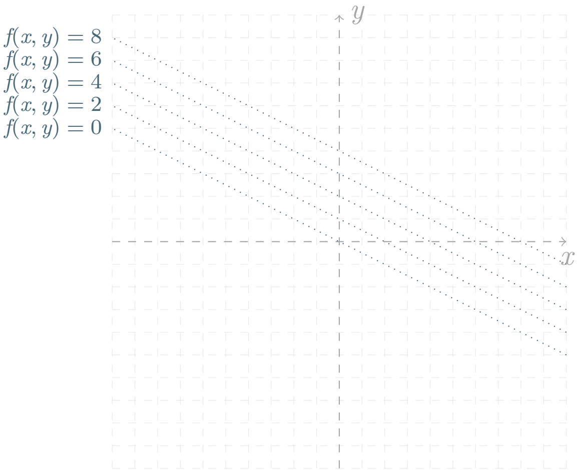

y
$f(x, y) = 0$
x


y
$f(x, y) = 2$
$f(x, y) = 0$
x

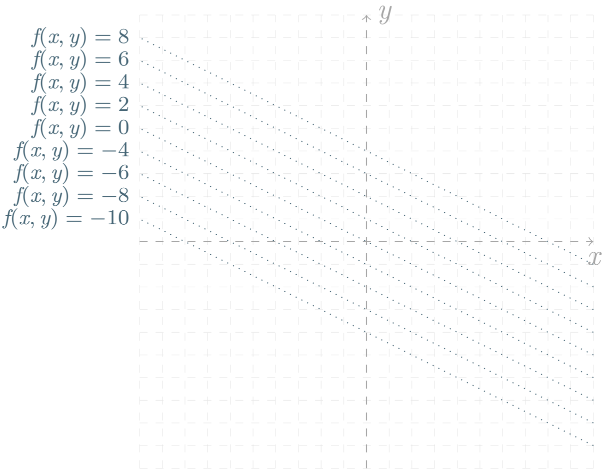


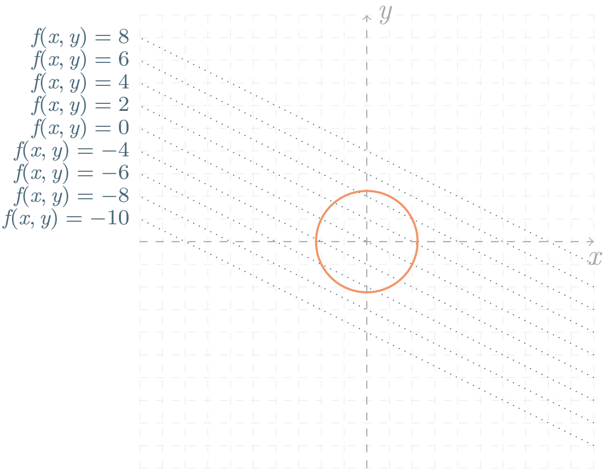
$f(x, y) = 4$
$f(x, y) = 2$
$f(x, y) = 0$
y
x


$f(x, y) = 6$
$f(x, y) = 4$
$f(x, y) = 2$
$f(x, y) = 0$
y
x

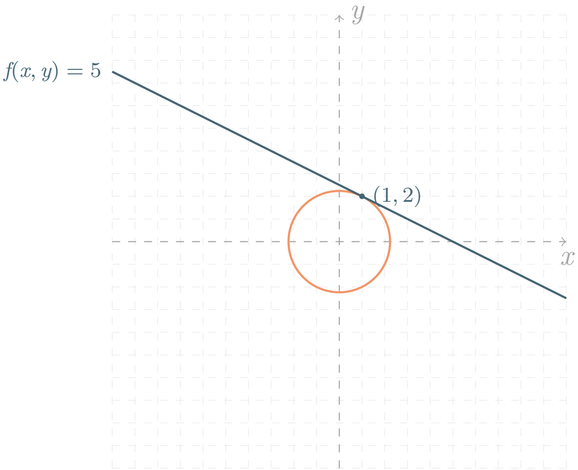


y
$f(x,y)=8$
$f(x,y)=6$
$f(x,y)=4$
$f(x,y)=2$
$f(x,y)=0$
x


$f(x, y) = 8$
$f(x, y) = 6$
$f(x, y) = 4$
$f(x, y) = 2$
$f(x, y) = 0$
$f(x, y) = -4$
$f(x, y) = -6$
$f(x, y) = -8$
$f(x, y) = -10$
y
x



$f(x, y) = 8$
$f(x, y) = 6$
$f(x, y) = 4$
$f(x, y) = 2$
$f(x, y) = 0$
$f(x, y) = -4$
$f(x, y) = -6$
$f(x, y) = -8$
$f(x, y) = -10$
y
x



$f(x,y)=5$
$(1,2)$
x
y

Take away:

- The direction of increase for our objective function must be perpendicular to the constraint.
- Perpendicular to constraint: direction of increase thinking of constraint as a function!
- The tangencies relative to objective function and constraint are parallel.

How do we operationalize this?

- We want to maximize $f : \mathbb{R}^n \to \mathbb{R}$
- subject to $g_j : \mathbb{R}^n \to \mathbb{R}$ are equal to 0, for $1 \leq j \leq k$:
- That is: $\max_{\boldsymbol{x}} f(\boldsymbol{x})$ such that $g_i(\boldsymbol{x}) = 0$.

We can use the method of Lagrange multipliers. The Lagrangian of the above problem is:

$$\mathscr{L}(\boldsymbol{x}) = f(\boldsymbol{x}) - \sum_{i \leq k} \lambda_i g_i(\boldsymbol{x}),$$

We then optimize $\mathscr{L}$ as before, including taking the derivative with respect to λ. We have $n + k$ first order conditions:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} - \sum_j \lambda_j \frac{\partial g_j(\boldsymbol{x})}{\partial x_i} = 0, \text{ for each } i \tag{$\mathscr{L}_i$}$$

$$g_j(\boldsymbol{x}) = 0, \text{ for each } j \tag{$\mathscr{L}_{\lambda_j}$}$$

We had $f(x, y) = x + 2y$, and $g(x, y) = x^2 + y^2$:

$$\mathscr{L}(x, y) = x + 2y - \lambda(x^2 + y^2)$$

The first order conditions are

$$1 - \lambda 2x = 0 \qquad (\mathscr{L}_x)$$

$$2 - \lambda 2y = 0 \qquad (\mathscr{L}_y)$$

$$x^2 + y^2 = 9 \qquad (\mathscr{L}_\lambda)$$

Substituting for λ we have $\frac{1}{y} = \frac{1}{2x}$ or that $y = 2x$. Plugging into $\mathscr{L}_\lambda$ we see that $5x^2 = 5$ or that $x = 1$, so $y = 2$.

Example

Maximize $f(x, y, z) = xyz$ subject to $x + y + z = 1$.

What is the Lagrangian?

Example

Maximize $f(x, y, z) = xyz$ subject to $x + y + z = 1$.

What is the Lagrangian?

$$\mathscr{L} = xyz - \lambda(x + y + z - 1)$$

Example

$\mathscr{L} = xyz - \lambda(x + y + z - 1)$

- We could take a monotone transformation to make easier.
- Lets maximize $\ln(f(x, y, z))$ instead.
- $\mathscr{L}' = \ln(x) + \ln(y) + \ln(z) - \lambda(x + y + z - 1)$

The FOCs are

$$x = \frac{1}{\lambda} \tag{$\mathscr{L}_x$}$$

$$y = \frac{1}{\lambda} \tag{$\mathscr{L}_y$}$$

$$z = \frac{1}{\lambda} \tag{$\mathscr{L}_z$}$$

$$x + y + z = 1 \tag{$\mathscr{L}_\lambda$}$$

So $x = y = z = \frac{1}{3}$.

Theorem. (The Constrained Envelope Theorem)

Let $f, g : \mathbb{R}^{n+k} \to \mathbb{R}$. We want to solve $\max_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{a})$ subject to $g(\boldsymbol{x}, \boldsymbol{a}) = 0$. Let $f^\star$ be the optimized value. Then so long as both partial derivatives exist, we have

$$\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{a}} - \lambda \frac{\partial g}{\partial \boldsymbol{a}}$$

What is λ?

- The marginal value of relaxing the constraint.
- Notice at the optimum:

$$\frac{\frac{\partial f}{\partial x_i}}{-\frac{\partial g}{\partial x_i}} = \frac{\frac{\partial f}{\partial x_j}}{-\frac{\partial g}{\partial x_j}} = \lambda$$

  - $\frac{\partial f}{\partial x_i}$ is the benefit to increasing x_i
  - $-\frac{\partial g}{\partial x_i}$ is the cost of increasing x_i.
- The ratio of cost to benefit for each good must be equal.

Inequality Constraints

What about constraints that are not binding?

- A consumer cannot spend more than her budget.
- A firm cannot pollute more than regulated maximum.
- A firm must produce enough to fill a contract.
- A government must meet a threshold tax revenue.

We want to allow $g_j(\boldsymbol{x}) = 0$ but also $g_j(\boldsymbol{x}) \leq 0$. If at the optimum

- $g_j(\boldsymbol{x}) = 0$ then the j^{th} constraint is binding.
- $g_j(\boldsymbol{x}) < 0$ then the j^{th} constraint is slack.

Again, we have the Lagrangian: $\mathscr{L}(\boldsymbol{x}) = f(\boldsymbol{x}) - \sum_j \mu_j g_i(\boldsymbol{x})$.

Theorem. (KKT)

If $\boldsymbol{x}^*$ is a maximizer (which meets some technical conditions[1]) of f subject to $g_1 \ldots g_k$, then there exists a $\mu^\star$ such that:

1. Stationarity, or FOC: $\nabla \mathscr{L}(\boldsymbol{x}^\star) = 0$
   - $\nabla f(\boldsymbol{x}^\star) = \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^\star)$
2. Positive multipliers: $\mu_j^\star \geq 0$
3. Complimentary slackness: $\mu_j^\star g_j(\boldsymbol{x}^\star) = 0$
4. Primal feasibility: $g_j(\boldsymbol{x}^\star) \leq 0$
5. Second order conditions are met.

[1] These conditions are call constraint qualification conditions which ensure the boundary of the feasible set is well behaved. If our constraints are linear, it is always met, so we will ignore such technicalities for this class.

There are two possibilities:

1. j is binding: $g_j(\boldsymbol{x}^*) = 0$
   - we should obtain the same solution as if we had used an equality constraint.
   - CS is met, no additional data on solution
2. j is slack: $g_j(\boldsymbol{x}^*) < 0$
   - Then CS implies $\mu_j = 0$
   - As if there is no constraint at all (we find the unconstrained maximum).

Example

- An agent can consume 2 goods, the amounts of which are denoted by x and y.
- She receives a utility of xy
- Prices are both 1; she has a budget of 100; cannot consume more than 40 units of good 1.

Example

She wants to maximize: $f(x, y) = xy$ subject to

$$\begin{aligned} x + y - 100 &\leq 0 \\ x - 40 &\leq 0 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

We have

$$\mathscr{L} = xy - \mu_1(x + y - 100) - \mu_2(x - 40) - \mu_x(-x) - \mu_y(-y)$$

Example

We have the first order conditions:

$$
\begin{aligned}
\mathscr{L}_x : \qquad & y - \mu_1 - \mu_2 - \mu_x = 0 \\
\mathscr{L}_y : \qquad & x - \mu_1 - \mu_y = 0 \\
\mathscr{L}_{\mu_1} : \qquad & \mu_1(x + y - 100) = 0 \\
\mathscr{L}_{\mu_2} : \qquad & \mu_2(x - 40) = 0 \\
\mathscr{L}_{\mu_x} : \qquad & \mu_x x = 0 \\
\mathscr{L}_{\mu_y} : \qquad & \mu_y y = 0
\end{aligned}
$$

We also have our non-negativity constraints given by 2 and 4 of the KKT theorem.

Example

1. $f(1,1) = 1 > 0 = f(x,0) = f(0,y)$ implies we can drop $\mathscr{L}_{\mu_x}$ and $\mathscr{L}_{\mu_x}$.
2. Now, what if $\mu_1 = 0$?
   - implies that $x = 0$ by $\mathscr{L}_y$. But we just argued $x > 0$ so we know that μ_1 must bind.
   - So we know $\mu_1 > 0$ implies $x = 100 - y$
3. what if $\mu_2 = 0$?
   - implies from $\mathscr{L}_x$ and $\mathscr{L}_y$ we would have $x = y = 50$
   - Violates the constraint: $\mu 2 > 0$.
4. So $x^\star = 40$. Hence, $y^\star = 60$.

Example

Consider maximizing $f(x, y) = \sum_{i=1}^{2} \ln(1 + \frac{x_i}{p_i})$ subject to $x_1 + x_2 \leq 1$ and $x_i \geq 0$ for each i. Assume each $p_i > 0$. The Lagrangian is

$$\mathscr{L} = \sum_{i=1}^{n} \ln(1 + \frac{x_i}{p_i}) - \mu_0(x_1 + x_2 - 1) - \sum_{i=1}^{n} \mu_i(-x_i).$$

Example

The KKT conditions:

$$
\begin{aligned}
&\frac{1}{p_i + x_i} = \mu_0 - \mu_i && (\mathscr{L}_{x_i})\\
&\mu_0(x_1 + x_2 - 1) = 0 && (CS_0)\\
&\mu_i x_i = 0 && (CS_i)\\
&\mu_0, \mu_i \geq 0 && (\geq)\\
&x_1 + x_2 \leq 1 && (FSB_0)\\
&x_i \geq 0 && (FSB_1)
\end{aligned}
$$

1. From $\mathscr{L}_{x_i}$: $\mu_0 - \mu_i > 0$ so $\mu_0 > 0$.
2. Therefore $x_1 = 1 - x_2$.
3. Can both x_i's be positive?
   - Implies by $\mathscr{L}_{x_1}$ and $\mathscr{L}_{x_2}$: $p_1 + x_1 = \frac{1}{\mu_0} = p_2 + x_2$
   - Feasible if and only if $|p_1 - p_2| \leq 1$

- Otherwise: $x_i^\star = 1$ for lower cost p_i.

What if we want to minimize f subject to g?

- This is the same as maximizing $-f$ subject to g.
- Therefore FOC states:

Given f subject to $g \leq 0$

- If maximizing f:

$$\mathscr{L}(\boldsymbol{x}) = f(\boldsymbol{x}) - \sum_j \mu_j g_i(\boldsymbol{x})$$

- If minimizing f:

$$\mathscr{L}(\boldsymbol{x}) = -f(\boldsymbol{x}) - \sum_j \mu_j g_i(\boldsymbol{x})$$

- The FOC therefore states:

$$-\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*)$$