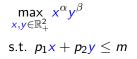
# Some More Math for Economists: Constrained Optimization

Spring 2023

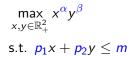
#### Optimization

In economic modeling, we explicitly model choices. This done through optimization.

$$\max_{\substack{x,y \in \mathbb{R}^2_+}} x^{\alpha} y^{\beta}$$
  
s.t.  $p_1 x + p_2 y \le m$ 



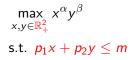
#### The endogenous variables



- The endogenous variables
- The exogenous variables

$$\max_{\substack{x,y \in \mathbb{R}^2_+}} \frac{x^{\alpha} y^{\beta}}{s.t. \ p_1 x + p_2 y \le m}$$
s.t. p\_1x + p\_2y  $\le m$ 

- The endogenous variables
- The exogenous variables
- ► The objective



- The endogenous variables
- The exogenous variables
- ► The objective
- The constraints (the feasible set)

#### Some Examples

Profit Maximization:

$$\max_{q\in\mathbb{R}_+}qP(q)-c(q)$$

#### Some Examples

Utility maximization:

 $\max_{x \in \mathbb{R}^n_+} u(x)$ <br/>s.t.  $p \cdot x \le m$ 

#### Some Examples

Consumption/Savings:

$$\max_{c,s} \sum_{t=0}^{T} \delta^{t} u(c_{t})$$
  
s.t.  $c_{t} + s_{t} = (1+r)s_{t-1} + w$   
 $s_{T} = s_{-1} = 0$ 

or in continuous time

$$\max \int_{t=0}^{T} e^{-rt} u(c_t)$$
  
s.t.  $\dot{s}_t = rs_t + w - c_t$   
 $s_0 = s_T = 0$ 

## Optimization

In this course we're going to develop some basic tools for non-linear optimization

- Our focus is primarily theoretical, we'll develop the tools we need to establish theoretical results in the subsequent economics courses.
- This course only scratches the surface. I leave work on numerical tools, stochastic optimal control, etc. to other courses.

# A very quick review

Existence:

- Knowing that a problem has a maximum is very important.
- The extreme value theorem tells us that any continuous function with a compact domain has a maximum/minimum.
- ► This result is very general.
- ► For some of our problems, its application is a bit subtle. I'm I'm mostly going to ignore questions of existence in this course.

Unconstrained optimization:

- First order conditions:  $\nabla f(x) = 0$  at an interior max.
- Second order conditions: D<sup>2</sup>f(x) is negative semidefinite at a max. ∇f(x) = 0 and D<sup>2</sup>f(x) negative definite are sufficient for a maximum.

Remember the consumer problem:

 $\max u(x_1, x_2)$ <br/>s.t.  $p_1 x_1 + p_2 x_2 = m$ 

Suppose I move x a little bit to  $(x_1 + dx_1, x_2 + dx_2)$ . What has to hold at a max?

First, we need to make this tiny change while maintaining the constraint, i.e.:

$$p_1dx_1+p_2dx_2=0$$

and we know if these changes are small

$$u(x + dx) - u(x) \approx dx_1 u_1(x_1, x_2) + dx_2 u_2(x_1, x_2).$$

Therefore at a max, for any feasible change

$$dx_1u_1(x_1, x_2) + dx_2u_2(x_1, x_2) \approx 0$$

We know that, at a max

$$p_1dx_1+p_2dx_2=0$$

 $\mathsf{and}$ 

$$dx_1u_1(x_1, x_2) + dx_2u_2(x_1, x_2) = 0$$

We know that, at a max

$$p_1dx_1+p_2dx_2=0$$

 $\mathsf{and}$ 

$$dx_1u_1(x_1, x_2) + dx_2u_2(x_1, x_2) = 0$$

Combining these:

$$dx_1u_1(x_1, x_2) = \frac{p_1}{p_2}dx_1u_2(x_1, x_2)$$
$$\frac{u_1(x_1, x_2)}{p_1} = \frac{u_2(x_1, x_2)}{p_2}$$

We know that, at a max

$$p_1dx_1+p_2dx_2=0$$

and

$$dx_1u_1(x_1, x_2) + dx_2u_2(x_1, x_2) = 0$$

Combining these:

$$dx_1 u_1(x_1, x_2) = \frac{p_1}{p_2} dx_1 u_2(x_1, x_2)$$
$$\frac{u_1(x_1, x_2)}{p_1} = \frac{u_2(x_1, x_2)}{p_2}$$

So, there exists a constant  $\lambda$  s.t.

$$\frac{u_1(x_1, x_2)}{p_1} = \frac{u_2(x_1, x_2)}{p_2} = \lambda$$

Which gives us the familiar condition

$$\nabla u(x) = \lambda(p_1, p_2)$$

We did two things here:

- We used the constraint to identify how variations in x<sub>2</sub> implicitly are determined by the perturbation we make to x<sub>1</sub>.
- ► We used this implicit function to make "first order conditions" Let's make things precise

## Lagrange Multipliers

#### Theorem

Let  $f : \mathbb{R}^n \to \mathbb{R}$  and  $g : \mathbb{R}^n \to \mathbb{R}^m$ . Let  $x^*$  be a solution to

 $\max f(x) \ s.t. \ g(x) = 0$ 

and suppose  $Dg(x^*)$  has rank m. Then there exists a unique Lagrange multiplier  $\lambda \in \mathbb{R}^m$  such that  $Df(x^*) = \lambda' Dg(x^*)$ .

# Lagrange Multipliers

Geometrically, what does it mean for such a  $\lambda$  to exist?

#### Example

Consider the consumer problem

$$\max_{\substack{x,y \in \mathbb{R}_+ \\ \text{s.t. } p_1 x + p_2 y = m}} x^{1/2} y^{1/2}$$

The Lagrange multiplier theorem gives us "FOCs"

$$\frac{1}{2}x^{-1/2}y^{1/2} = \lambda p_1$$
$$\frac{1}{2}x^{1/2}y^{-1/2} = \lambda p_2$$

Turns our maximization problem into a system of three non-linear equations with three unknowns

#### Example

Combining the FOCs, we get

$$\frac{1}{2p_1}x^{-1/2}y^{1/2} = \frac{1}{2p_2}x^{1/2}y^{-1/2}$$
$$p_2y = p_1x$$

Plugging this into the constraint, we get

$$x = \frac{m}{2p_1}, y = \frac{m}{2p_2}$$

#### Example

We can also apply this theorem to problems with more than one constraint. Consider:

max 
$$4y - 2z$$
  
s.t.  $2x - y - z = 2$   
 $x^{2} + y^{2} = 1$ 

#### Lagrangian

Before we get to the proof, one more useful object. We can define the function

$$L(\lambda) = \max_{x \in X} f(x) - \lambda g(x).$$

This is called the Lagrangian. It's easy to see that the FOCs of the unconstrained maximization problem are our Lagrange multiplier conditions.

- The multiplier puts a "price" on how much we violate each constraint.
- If (x\*, λ\*) solved the constrained maximization problem, then if x\* solves f(x) − λ\*g(x) then L(λ\*) = f(x\*).

This is not obviously true, even though the FOCs are satisfied!

► For many problems we'll see, this is true.

# Lagrange Multipliers

Now let's prove the theorem. We need a tool from Math for Economists to show this

Theorem (Implicit Function Theorem)

Let f(x, y) be a continuously differentiable function  $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ . Fix a point (a, b) s.t. f(a, b) = 0. Suppose  $D_y f(a, b)$  is invertible. Then there exists an open set U containing a and a unique continuously differentiable function  $g : U \to \mathbb{R}^m$ such that g(a) = b and f(x, g(x)) = 0 on U. g satisfies the differential equation

$$Dg(x) = -(D_y f(x, g(x)))^{-1} D_x f(x, g(x)).$$

Consider the program

$$\max f(x,y) \text{ s.t. } g(x,y) = 0$$

where  $f : \mathbb{R}^n \to \mathbb{R}$ ,  $g : \mathbb{R}^n \to \mathbb{R}^m$ ,  $y \in \mathbb{R}^m$  and  $x \in \mathbb{R}^{n-m}$ .

If the conditions of the implicit function theorem hold, we can find an  $h : \mathbb{R}^{n-m} \to \mathbb{R}^m$  s.t. g(x, h(x)) = 0 in a nbhd of any maximum.

#### Lagrange Multipliers

So at a maximum,  $(x^*, y^*)$ ,  $x^*$  must solve

 $\max_{x\in U}f(x,h(x))$ 

where  $h(x^*) = y^*$ , g(x, h(x)) = 0 for all  $x \in U$  and h is differentiable.

FOCs:

$$D_x f(x, h(x)) + D_y f(x, h(x)) Dh(x) = 0$$

From the implicit function theorem we know

$$Dh(x) = -(D_y g(x, h(x)))^{-1} D_x g(x, h(x))$$
  
So if  $\lambda' = D_y f(x, h(x)) D_y g(x, h(x))^{-1}$  then  
$$D_x f(x^*, y^*) = \lambda' D_x g(x^*, y^*)$$

From the definition of λ we also have

$$D_y f(x^*, y^*) = \lambda' D_y g(x^*, y^*)$$

# Constraint Qualification

What did we need for this beyond the obvious (e.g. differentiability)?

- to use the implicit function theorem, Dyg(x\*, y\*) must be full rank.
- We have some flexibility here, it doesn't really matter which m components we called "y"
- It's hard to assume this away. I can formulate any set of constraints in a way that this condition is violated at every feasible point.

Constraint Qualification - An Example

Consider the consumer problem

max 
$$x^{1/2}y^{1/2}$$
  
s.t.  $p_1x + p_2y = m$ 

This satisfies constraint qualification at every feasible point and is solved by

$$x=\frac{m}{2p_1}, y=\frac{m}{2p_2}$$

#### Constraint Qualification - An Example

What if we instead tried to apply our tool to

$$\max x^{1/2} y^{1/2}$$
  
s.t.  $(p_1 x + p_2 y - m)^3 = 0$ 

This is the exact same problem. But now the FOCs are

$$\frac{1}{2}x^{-1/2}y^{1/2} = 3\lambda p_1(p_1x + p_2y - m)^2$$
$$\frac{1}{2}x^{1/2}y^{-1/2} = 3\lambda p_2(p_1x + p_2y - m)^2$$

which simplify to

$$\frac{1}{2}x^{-1/2}y^{1/2} = 0$$
$$\frac{1}{2}x^{1/2}y^{-1/2} = 0$$

which is clearly not satisfied at the maximum.

This example is a bit extreme, at every feasible point the constraint has 0 gradient.

- In practice, this is more manageable.
- To find candidate maxes we need to find
  - All points where the Lagrange multiplier conditions hold
  - All points where Rank  $Dg(x^*) \neq m$ .
- If a max exists, it's one of these points

Conceptually, there's no reason to require our consumer to spend all their money. The consumer problem should be

 $\max_{x\in\mathbb{R}^n_+}u(x)$ s.t.  $p\cdot x\leq m$ 

What can we do with this? It turns out, Lagrange multipliers still "work"

#### Non-negativity Constraints

Let's think about the simplest inequality constraints, constraints of the form

$$x_i \ge 0$$

Consider

 $\max u(x)$ <br/>s.t.  $x_1 \ge 0$ 

Let  $x^*$  be a max. Either  $x^* >> 0$  and optimality implies  $\nabla u(x) = 0$  or  $x_1^* = 0$ .

#### Non-negativity

Suppose  $x_1^* = 0$ ,  $x_i^* > 0$  for all  $i \neq 1$ . Then if I look at  $x^* + dx$  it must be that

$$u(x^*+dx)-u(x)\leq 0$$

for any dx small s.t.  $dx_1 \ge 0$ . This means approximately

$$\sum_{i=1}^n u_i(x^*) dx_i \leq 0$$

which gives  $u_i(x^*) = 0$  for all  $i \neq 1$  and

 $u_1(x^*) \leq 0$ 

#### Non-negativity

So we have two cases. Either

$$x_1^*=0$$
 and  $u_1(x^*)\leq 0,$   $u_i(x^*)=0$   $orall i
eq 1$ 

or

$$x_1^* > 0$$
 and  $\nabla u(x^*) = 0$ .

We can formulate these conditions using Lagrange multipliers. It must be that

$$\nabla u(x) = \lambda(-1, 0, 0, \ldots)$$

 $\mathsf{and}$ 

 $\lambda \geq \mathbf{0}$ 

and

$$\lambda x_1 = 0$$

at any maximum.

# **KKT** Conditions

Back to the consumer problem. Let's add a new variable s (for "slack"), and instead solve

$$\max_{x \in \mathbb{R}^{n}_{+}, s \in \mathbb{R}} u(x)$$
  
s.t.  $p \cdot x + s = m$   
 $s \ge 0$ 

Ignoring non-negativity of the x's, this gives us Lagrange multiplier condition

$$\nabla u(x) = \lambda p$$

and the additional conditions

$$egin{aligned} 0 &= \lambda - \mu \ \mu(-s) &= 0 \ \mu &\geq 0 \end{aligned}$$

#### **KKT** Conditions

Now let's get rid of the auxiliary variables by noting:

$$\lambda = \mu$$

and

$$s = m - p \cdot x$$
.

We end up with the following conditions:

$$abla u(x) = \lambda p$$
 $\lambda(m - p \cdot x) = 0$ 
 $\lambda \ge 0$ 

# KKT conditions

In general

#### Theorem (Karush-Kuhn-Tucker Conditions)

Let  $f : \mathbb{R}^n \to \mathbb{R}$  and  $g : \mathbb{R}^n \to \mathbb{R}^m$ , differentiable. Suppose  $x^*$  solves

$$\max f(x) \ s.t \ g(x) \leq 0$$

and rank  $Dg^*(x^*) = m^*$  where  $g^*$  is the vector of binding constraints and  $m^*$  is the number of binding constraints. Then there exists a  $\lambda \in \mathbb{R}^m$  such that

$$\begin{aligned} Df(x^*) &= \lambda' Dg(x^*) \\ g(x^*)_i &\leq 0 \text{ for all } i \in \{1, 2, \dots m\} \\ \lambda_i g(x^*)_i &= 0 \text{ for all } i \in \{1, 2, \dots m\} \\ \lambda_i &\geq 0 \text{ for all } i \in \{1, 2, \dots m\} \end{aligned}$$

# **KKT** Conditions

#### We have

$$\begin{aligned} Df(x^*) &= \lambda' Dg(x^*) \\ g(x^*)_i &\leq 0 \text{ for all } i \in \{1, 2, \dots m\} \\ \lambda_i g(x^*)_i &= 0 \text{ for all } i \in \{1, 2, \dots m\} \\ \lambda_i &\geq 0 \text{ for all } i \in \{1, 2, \dots m\} \end{aligned}$$

- The first three conditions we could have essentially reached mechanically from the equality constraint result.
- The fourth is new. It is a consequence of the inequality constraint.
- For minimization problems, the multipliers must be negative.

#### Example - KKT

# $\max xy$ <br/>s.t. $x^2 + y^2 \le 1$

# Example - KKT

$$\max xyz + z$$
  
s.t.  $x^2 + y^2 + z \le 6$   
 $x, y, z \ge 0$ 

# Example - KKT

$$\max_{\substack{x,y \in \mathbb{R}_+ \\ \text{s.t } y - (1-x)^3 \le 0}}^{\max x}$$

### Example - Envy

There are two consumers, each of whom are jealous of the others consumption, captured by utility function

$$u_i(x_i, x_j) = x_i - K x_j^2$$

and there are in total X units of consumption in the economy. A social planner solves

$$\max_{x_1,x_2 \in \mathbb{R}_+} u_1(x_1,x_2) + u_2(x_2,x_1) ext{ s.t. } x_1 + x_2 \leq X$$

As a function of K, what is the efficient allocation? Does the constraint bind?

#### Lagrangian

Let's think about the Lagrangian again

$$L(\lambda) = \max_{x \in X} f(x) - \lambda g(x)$$

Let  $(x^*, \lambda^*)$  be a maximum and a corresponding multiplier that satisfies the KKT conditions. Then for any  $\lambda \ge 0$ 

$$egin{aligned} \mathcal{L}(\lambda) &= \max_{x \in X} f(x) - \lambda g(x) \ &\geq f(x^*) - \lambda g(x^*) \ &\geq f(x^*) - \lambda^* g(x^*) = f(x^*) \end{aligned}$$

where the third line comes from complementary slackness. Therefore

$$\min_{\lambda \ge 0} L(\lambda) \ge f(x^*)$$

This is called duality. It's reasonable to expect, but not obvious that this  $\geq$  is an = for "nice" problems.

# **KKT** Conditions

We're left with a few loose ends we'd like to tie up. The KKT conditions aren't quite necessary or sufficient.

- (Necessity) The KKT conditions don't hold at maxima where the derivative matrix of the binding constraints is not full rank.
- ► (Sufficiency) If λ > 0, we know the point can't be a min. Is that enough to tell us it's a max?

It turns out that we can make economically meaningful assumptions that also ensure these conditions are both necessary and sufficient.