Constrained Optimization

Econ 6105, Fall 2024

Prof. Josh Abel

(SB Chapters 18, 19.1, 19.3, 16.3)

Optimizing Under Constraints

Ubiquitous problem in economics:

- Agent is "rational," i.e. chooses best available option
- Agent is constrained, i.e. cannot have everything she wants

In math terms, this is "constrained optimization."

Canonical problem:

- Agent gets utility from two goods, given by $u(x_1, x_2)$.
- Agent has income I and faces constant prices, p_1 and p_2
  - Normalize $p_1 = 1$, denote $p \equiv p_2$
- We'll look at the more general problem later, but the 2-good setting is very instructive (and common)

Two typical approaches:

1. Substitution
2. Lagrangian

Substitution

Agent's problem: maximize $u(x_1, x_2)$ s.t. $x_1 + p \cdot x_2 \leq I$.
Assume $u_1, u_2 > 0$, so the agent will spend all income. Constraint:

$$x_1 + p \cdot x_2 = I \tag{1}$$

Can formulate *unconstrained* problem via substitution:

$$\max_{x_2} u(I - p \cdot x_2, x_2) \tag{2}$$

Find local extrema: set $\frac{du}{dx_2} = 0$

$$\frac{du}{dx_2} = -p \cdot u_1(I - p \cdot x_2^*, x_2^*) + u_2(I - p \cdot x_2^*, x_2^*) = 0 \tag{3}$$

or more succinctly:

$$\frac{du}{dx_2} = -p \cdot u_1 + u_2 = 0 \tag{4}$$

This is known as a "First-Order Condition (FOC)": any local max of a differentiable function will obey this

Substitution (2)

FOC implicitly characterizes x_2^*, so we have two equations and 2 unknowns:

1. $p \cdot u_1(x_1^*, x_2^*) = u_2(x_1^*, x_2^*)$
2. $x_1^* + p \cdot x_2^* = I$

Technically, need to make sure second derivative/Hessian is negative (definite): we'll get to that later.

Other than that, we're kind of done. Can:

1. Mine implicit solution for insight
    - FOC: $u_2/u_1 = p$ (Marginal Rate of Substitution equals price ratio)
2. Make parametric assumptions to get explicit solution
    - If $u = x_1^\alpha \cdot x_2^{1-\alpha}$, then $x_1^* = \alpha \cdot I$ and $x_2^* = (1-\alpha) \cdot I/p$

Lagrangian

Lagrangian approach is more involved but more powerful. We define a new object L and a constant λ with:

$$L(x_1, x_2, \lambda) = \max_{x_1, x_2, \lambda} u(x_1, x_2) + \lambda \cdot (I - x_1 - p \cdot x_2) \qquad (5)$$

For this to have a local max at some $(x_1^*, x_2^*, \lambda^*)$, 1 of 2 things must be true:

1. Either: $I = x_1^* + p \cdot x_2^*$;
2. Or: $\lambda^* = 0$

If $I \neq x_1^* + p \cdot x_2^*$, then we can get L to go to ∞ with $\lambda \to \infty$ (or $-\infty$), so it would not be a local max.
The Lagrangian is essentially a trick to get our tools from unconstrained problems to carry over to a constrained problem

- The possibility of $\lambda^* = 0$ is a complication we will discuss later

Lagrangian Cookbook

FOCs ($\nabla L(x_1^*, x_2^*, \lambda^*) = 0$):

1. $\frac{\partial L}{\partial x_1} = u_1(x_1^*, x_2^*) - \lambda^* = 0$
2. $\frac{\partial L}{\partial x_2} = u_2(x_1^*, x_2^*) - p \cdot \lambda^* = 0$
3. $\frac{\partial L}{\partial \lambda} = I - x_1^* - p \cdot x_2^* = 0$

This is 3 equations in 3 unknowns. Can rearrange as:

1. $p \cdot u_1(x_1^*, x_2^*) = u_2(x_1^*, x_2^*)$
2. $x_1^* + p \cdot x_2^* = I$
    - 1-2 are exactly the same as the Substitution approach
3. $\lambda^* = u_1(x_1^*, x_2^*) = u_2(x_1^*, x_2^*)/p$

Understanding the Solution: Univariate Intuition

Consider the nearly-trivial univariate problem of maximizing $f(x) = -(2-x)^2$, constrained by $x \leq 1.5$

Solution is obviously to get as close to $x = 2$ as possible.

- Given the constraint, $x^* = 1.5$.

Applying the cookbook:

$$L(x, \lambda) = -(2-x)^2 + \lambda \cdot (1.5 - x)$$

FOCs:

1. wrt x: $2 \cdot (2-x) = \lambda^*$
2. wrt λ: $1.5 = x^*$

λ^* reveals how much better we could do if the constraint were eased.

- $\lambda^* = 1 = \frac{df}{dx}(1.5)$ – slope of f when we were forced to stop at $x = 1.5$
- "Shadow price" of the constraint

Understanding the Solution: Back to the Multivariate Problem

$$\lambda^* = u_1(x_1^*, x_2^*) = u_2(x_1^*, x_2^*)/p$$

λ^* is the "marginal utility of income"
A small change in income of dI will increase utility by $\lambda^* \cdot dI$

1. Or $u_1 \cdot dI$
2. Or $u_2/p \cdot dI$

What is the economic intuition for why $u_1 = u_2/p$?
Recall: $\frac{dx_2}{dx_1}|_{dU=0} = \frac{MU_1}{MU_2}$
So: $\frac{dx_2}{dx_1}|_{dU=0} = p$.

- Slope of constraint (p) equals slope of objective's contour map.

Many Goods and Many (Equality) Constraints

These ideas generalize with many goods and constraints
Let $f : R^n \rightarrow R^1$ be a differentiable objective function and $h_1, ..., h_m : R^n \rightarrow R^1$ be differentiable equality constraint functions.

- I.e. we want to find $x \in R^n$ that maximizes f, where $h_1(x) = a_1$,...$h_m(x) = a_m$.

For, x^*, a local extremum in the constrained subset of R^n, there exist λ_1^*,...,λ_m^* that satisfy:

$$\frac{\partial L}{\partial x_i} = 0 \text{ for } i = 1, ..., n \quad (6)$$

$$\frac{\partial L}{\partial \lambda_j} = 0 \text{ for } j = 1, ..., m \quad (7)$$

for the following Lagrangian:

$$L(x, \lambda) \equiv f(x) + \lambda_1 \cdot (a_1 - h_1(x)) + ... + \lambda_m \cdot (a_m - h_m(x)) \quad (8)$$

Practice Problem

Maximize $f(x, y, z) = x^{1/2} + y^{1/2} + z^{1/2}$ such that $x + y + z = 17$ and $x \cdot y = 16$.
Lagrangian:

Practice Problem

Maximize $f(x, y, z) = x^{1/2} + y^{1/2} + z^{1/2}$ such that $x + y + z = 17$ and $x \cdot y = 16$.
Lagrangian:

$$L(x, y, z, \lambda_1, \lambda_2) = x^{1/2} + y^{1/2} + z^{1/2} + \lambda_1 \cdot (17 - x - y - z) + \lambda_2 \cdot (16 - x \cdot y)$$

FOCs:

Practice Problem

Maximize $f(x, y, z) = x^{1/2} + y^{1/2} + z^{1/2}$ such that $x + y + z = 17$ and $x \cdot y = 16$.
Lagrangian:

$$L(x, y, z, \lambda_1, \lambda_2) = x^{1/2} + y^{1/2} + z^{1/2} + \lambda_1 \cdot (17 - x - y - z) + \lambda_2 \cdot (16 - x \cdot y)$$

FOCs:

1. $\frac{1}{2} \cdot x^{-1/2} = \lambda_1 + \lambda_2 \cdot y$
2. $\frac{1}{2} \cdot y^{-1/2} = \lambda_1 + \lambda_2 \cdot x$
3. $\frac{1}{2} \cdot z^{-1/2} = \lambda_1$
4. $17 = x + y + z$
5. $16 = x \cdot y$

(The “*”s are suppressed for readability.)

Practice Problem: Solution

FOCs 1 and 2: $x^* = y^*$
FOC 5: $x^* = y^* = 4$
So, FOC 4: $z^* = 9$
FOC 3: $\lambda_1^* = 1/6$
FOC 1 (or 2): $\lambda_2^* = 1/48$
$f(x^*, y^*, z^*) = 7$
Suppose we changed the first constraint to $x + y + z = 18$. What do you think $f(x^*, y^*, z^*)$ would be?
Suppose we changed the second constraint to $x \cdot y = 17$. What do you think $f(x^*, y^*, z^*)$ would be?

Practice Problem: Solution

FOCs 1 and 2: $x^* = y^*$
FOC 5: $x^* = y^* = 4$
So, FOC 4: $z^* = 9$
FOC 3: $\lambda_1^* = 1/6$
FOC 1 (or 2): $\lambda_2^* = 1/48$
$f(x^*, y^*, z^*) = 7$
Suppose we changed the first constraint to $x + y + z = 18$. What do you think $f(x^*, y^*, z^*)$ would be?
$\approx 7 + 1/6$
Suppose we changed the second constraint to $x \cdot y = 17$. What do you think $f(x^*, y^*, z^*)$ would be?
$\approx 7 + 1/48$
You can confirm these on your own.

Shortcomings of the Approach So Far

Return to the univariate problem but change the constraint:

$$L(x, \lambda) = -(2 - x)^2 + \lambda \cdot (2.5 - x)$$

FOCs:

1. wrt λ: $2.5 = x^*$
2. wrt x: $2 \cdot (2 - x) = \lambda^* \rightarrow \lambda^* = -1$

We know this is wrong; optimal choice is $x = 2 \neq 2.5$. What happened?

Shortcomings of the Approach So Far

Return to the univariate problem but change the constraint:

$$L(x, \lambda) = -(2 - x)^2 + \lambda \cdot (2.5 - x)$$

FOCs:

1. wrt λ: $2.5 = x^*$
2. wrt x: $2 \cdot (2 - x) = \lambda^* \rightarrow \lambda^* = -1$

We know this is wrong; optimal choice is $x = 2 \neq 2.5$. What happened? FOC wrt λ imposes that the constraint holds *with equality*.

- It correctly ruled out any possibility with $x > 2.5$
- But it also ignored any possibility with $x < 2.5$

To allow for the possibility that the constraint won't bind (i.e. will be "slack"), we need a more involved cookbook.

A Simple Problem with An Inequality Constraint

Use same Lagrangian as before:

$$L(x, \lambda) = -(2 - x)^2 + \lambda \cdot (2.5 - x)$$

Still take a FOC with respect to x:

1. $2 \cdot (2 - x^*) - \lambda^* = 0$

But FOC wrt λ is replaced with "complementary slackness conditions:"

1. $\lambda^* \cdot (2.5 - x^*) = 0$
2. $\lambda^* \geq 0$
3. $2.5 - x^* \geq 0$

Comp. Slack #1: "Either the constraint binds ($x^* = 2.5$) or $\lambda^* = 0$."

1. $x^* = 2.5$: $\lambda^* = -1$, which violates Comp. Slack #2!
   - Note, $f(2.5) = -0.25$
2. $\lambda^* = 0$: $x^* = 2$ (from FOC)
   - Correct answer: we've maximized $f(x)$ at $x = 2$, and we obey all conditions
   - $\lambda^* = 0$ means we do not benefit from loosening the constraint because it is already irrelevant/slack. Constraint disappears from calculations.

Two-Good Problem with Inequality Constraint

Maximize $u(x_1, x_2)$ s.t. $x_1 + p \cdot x_2 \leq I$.

$$L(x_1, x_2, \lambda) = \max_{x_1, x_2, \lambda} u(x_1, x_2) + \lambda \cdot (I - x_1 - p \cdot x_2)$$

Solution obeys:

Two-Good Problem with Inequality Constraint

Maximize $u(x_1, x_2)$ s.t. $x_1 + p \cdot x_2 \leq I$.

$$L(x_1, x_2, \lambda) = \max_{x_1, x_2, \lambda} u(x_1, x_2) + \lambda \cdot (I - x_1 - p \cdot x_2)$$

Solution obeys:

1. $u_1 = \lambda^*$
2. $u_2 = p \cdot \lambda^*$
3. $\lambda^* \cdot (I - x_1^* - p \cdot x_2^*) = 0$
4. $\lambda^* \geq 0$
5. $x_1^* + p \cdot x_2^* \leq I$

2 possibilities:

Two-Good Problem with Inequality Constraint

Maximize $u(x_1, x_2)$ s.t. $x_1 + p \cdot x_2 \leq I$.

$$L(x_1, x_2, \lambda) = \max_{x_1, x_2, \lambda} u(x_1, x_2) + \lambda \cdot (I - x_1 - p \cdot x_2)$$

Solution obeys:

1. $u_1 = \lambda^*$
2. $u_2 = p \cdot \lambda^*$
3. $\lambda^* \cdot (I - x_1^* - p \cdot x_2^*) = 0$
4. $\lambda^* \geq 0$
5. $x_1^* + p \cdot x_2^* \leq I$

2 possibilities:

1. $\lambda^* = 0$ (slack constraint: some income unspent)
   - Implies $u_1 = u_2 = 0$
2. $x_1^* + p \cdot x_2^* = I$
   - Same solution we found previously $(u_1 = u_2/p)$

Inequality Constraints in Practice

In most economic settings, the constraint is an inequality

- E.g. "Spend at or below your income."

However, in most economic models, the constraint will bind.

- We typically assume people will always want more.
  - Mathematically, $u_1, ..., u_n > 0$.

So in practice, we typically do not bother with the complementary slackness conditions.

- Say something like, "Due to positive marginal utility, the constraint will bind."
- Then, you can use the simpler cookbook for equality constraints, just setting $\frac{\partial L}{\partial x} = 0$ and $\frac{\partial L}{\partial \lambda} = 0$.

But if you're ever in a non-standard setting where a constraint might not bind, you need to go through the full process with the complementary slackness conditions!

Constrained Local Maxima In General

SB Theorem 18.5

Let x^* be a local maximum of $f(x) : R^n \to R^1$, a differentiable objective function, on the set of x that respect the following constraints:

$g_1(x) \leq b_1, ..., g_K(x) \leq b_K$

$h_1(x) = c_1, ..., h_M(x) = c_M$.

Assume all g and h functions are differentiable. Then, with a Lagranian defined as:

$$L(x, \lambda, \mu) \equiv f(x) + \sum_{k=1}^{K} \lambda_k \cdot (b_k - g_k(x)) + \sum_{m=1}^{M} \mu_m \cdot (c_m - h_m(x)),$$

there exist $\lambda_1^*,...,\lambda_K^*$, and $\mu_1^*,...,\mu_M^*$ such that:

1. $\frac{\partial L(x^*,\lambda^*,\mu^*)}{\partial x_1} = 0,...,\frac{\partial L}{\partial x_n} = 0$
2. $h_1(x^*) = c_1,...,h_M(x^*) = c_M$
3. $\lambda_1^* \cdot (b_1 - g_1(x^*)) = 0,...,\lambda_K^* \cdot (b_K - g_K(x^*)) = 0$
4. $\lambda_1 \geq 0,...,\lambda_K \geq 0$
5. $g_1(x^*) \leq b_1,...,g_K(x^*) \leq b_K$

Constrained Local Minima In General

Let x^* be a local maximum of $f(x) : R^n \rightarrow R^1$, a differentiable objective function, on the set of x that respect the following constraints:
$g_1(x) \leq b_1, ..., g_K(x) \leq b_K$
$h_1(x) = c_1, ..., h_M(x) = c_M$.
Assume all g and h functions are differentiable. Then, with a Lagranian defined as:

$$L(x, \lambda, \mu) \equiv f(x) + \sum_{k=1}^{K} \lambda_k \cdot (b_k - g_k(x)) + \sum_{m=1}^{M} \mu_m \cdot (c_m - h_m(x)),$$

there exist $\lambda_1^*,...,\lambda_K^*$, and $\mu_1^*,...,\mu_M^*$ such that:

1. $\frac{\partial L(x^*,\lambda^*,\mu^*)}{\partial x_1} = 0,...,\frac{\partial L}{\partial x_n} = 0$
2. $h_1(x^*) = c_1,...,h_M(x^*) = c_M$
3. $\lambda_1^* \cdot (b_1 - g_1(x^*)) = 0,...,\lambda_K^* \cdot (b_K - g_K(x^*)) = 0$
4. $\lambda_1 \leq 0,...,\lambda_K \leq 0$
5. $g_1(x^*) \leq b_1,...,g_K(x^*) \leq b_K$

Second-Order Condition

A local maximum of a problem with equality constraints should obey the FOCs we've focused on so far:

- $\frac{\partial L}{\partial x_i} = 0$, $\frac{\partial L}{\partial \lambda_j} = 0$

But even if some $\mathbf{x}$ satisfies the FOCs, it may not be a local maximum. It could be:

- A local minimum
- Neither a max or a min

For *unconstrained* optimization, we saw that a critical point ($\nabla f(\mathbf{x}) = 0$) is a maximum if its Hessian, $H(\mathbf{x})$, is negative definite.

- As a Hessian is the multivariate second derivative, this is called a "Second-Order Condition (SOC)"

Things are a bit harder in *constrained* maximization, but it still comes down to the negative definiteness of a Hessian.

We will start with a derivation with a 2-dimensional $\mathbf{x}$ with a linear constraint, but the ideas hold in more general settings.

$f : R^2 \rightarrow R^1$, Single Linear Constraint SOC Derivation

Maximize $f(x_1, x_2)$ s.t. $x_2 = \frac{Y - x_1}{p} \equiv \phi(x_1)$.

- Define $g(x_1) \equiv f(x_1, \phi(x_1))$

Now have unconstrained problem: maximize g. So need to find x_1^* s.t. $g'(x_1^*) = 0$ and $g''(x_1^*) < 0$.
Chain Rule: $\frac{dg}{dx_1}(x_1^*) = \frac{\partial f}{\partial x_1}(x_1^*, \phi(x_1^*)) + \frac{\partial f}{\partial x_2}(x_1^*, \phi(x_1^*)) \cdot \frac{d\phi}{dx_1}(x_1^*)$

- For concision, will say that FOC is $g' = f_1 + f_2 \cdot \phi' = 0$

Use Chain Rule again to get second derivative:
$g'' = f_{11} + f_{12} \cdot \phi' + (f_{21} + f_{22} \cdot \phi') \cdot \phi'$

- SOC: $f_{11} + 2 \cdot f_{12} \cdot \phi' + f_{22} \cdot (\phi')^2 < 0$

Bordered Hessian

Write constraint as $h(x_1, x_2) = c$:

- $h(x_1, x_2) = x_1 + p \cdot x_2 = Y$

Lagrangian:

$$L(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda(Y - h(x_1, x_2))$$

Lagranian's Hessian, called "Bordered Hessian:"

$$\bar{H}(\mathbf{x}, \lambda) \equiv \begin{bmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial x_1} & \frac{\partial^2 L}{\partial \lambda \partial x_2} \\ \frac{\partial^2 L}{\partial \lambda \partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \frac{\partial^2 L}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 0 & -h_1 & -h_2 \\ -h_1 & f_{11} & f_{12} \\ -h_2 & f_{12} & f_{22} \end{bmatrix} = \begin{bmatrix} 0 & -1 & -p \\ -1 & f_{11} & f_{12} \\ -p & f_{12} & f_{22} \end{bmatrix}$$

- Top-left is 0
- Bottom-right is Hessian of $f(x_1, x_2)$
- Upper border is gradient of constraint
- Left border is also gradient of constraint

Bordered Hessian and SOC

Determinant of Bordered Hessian:

$$det \begin{pmatrix} 0 & -1 & -p \\ -1 & f_{11} & f_{12} \\ -p & f_{12} & f_{22} \end{pmatrix} = 0 - -1 \cdot (-f_{22} - -p \cdot f_{12}) + -p \cdot (-f_{12} - -p \cdot f_{11})$$

$= -p^2 \cdot f_{11} + 2 \cdot p \cdot f_{12} - f_{22}$.
So $det(H(\mathbf{x}, \lambda) > 0 \iff f_{11} - \frac{2}{p} \cdot f_{12} + f_{22} \cdot \frac{1}{p^2} < 0$
Recall our SOC from earlier:

- $f_{11} + 2 \cdot f_{12} \cdot \phi' + f_{22} \cdot (\phi')^2 < 0$, where $\phi'(x) = -1/p$
- SOC holds $\iff f_{11} - \frac{2}{p} \cdot f_{12} + f_{22} \cdot \frac{1}{p^2} < 0$

In other words, SOC holds (i.e. we found a max) when the determinant of the Bordered Hessian is positive!

Extending to Many xs and Many Constraints

More generally, if you have N goods and K constraints, the Bordered Hessian should be $(N+K)x(N+K)$, with the same 4 regions (see e.g. SB Chapter 19, Equation 15):

- Top-left is a KxK matrix of 0
- Bottom-right is a NxN Hessian of $f(x_1, x_2)$
- Upper right (next to the 0s, above the Hessian) is a KxN matrix, where the top row is the gradient of the first constraint, etc.
- Bottom left (below to the 0s, next to the Hessian) is a NxK matrix, where the left column is the gradient of the first constraint, etc.

The SOC holds if the determinant of the Bordered Hessian has the same sign as $(-1)^N$ and the determinants of the largest $N-K$ principal submatrices have alternating signs.

- So in the $N=2$, $K=1$ case we did, we only had to check one determinant. In larger problems, there will be more computation.

SOC for a Minimization Problem

For a Bordered Hessian as described on the previous slide, the SOC for a minimization problem holds if the determinant of the Bordered Hessian and all $N - K$ of its largest principal submatrices all have the same signs as $(-1)^N$.

Concavity and Optimization

A concave function, f, is one such that for all $t \in [0,1]$:

$$f(t \cdot \mathbf{x} + (1-t) \cdot \mathbf{y}) > t \cdot f(\mathbf{x}) + (1-t) \cdot f(\mathbf{y}) \tag{9}$$

In the univariate case, this amounts to having a negative second derivative: $f''(x) < 0$.
This intuition carries over to the multivariate case: a multivariate function f is concave if and only if its Hessian is negative definite.

- Consider $\mathbf{x}$ and $\mathbf{y}$ in f's domain and define $g(t) \equiv f(t \cdot \mathbf{x} + (1-t) \cdot \mathbf{y})$.
- Can show $g''(t) = (\mathbf{x} - \mathbf{y})^T \cdot H \cdot (\mathbf{x} - \mathbf{y})$ (see SB, p. 514)
- So $g''(t) < 0$ for all $\mathbf{x}$, $\mathbf{y}$, and t precisely when H is negative definite.

So if you know your function is concave, you do not need to worry about SOC: it will be satisfied.

Quasiconcavity

A quasiconcave function, f, is one such that for all $t \in [0, 1]$:

$$f(t \cdot \mathbf{x} + (1 - t) \cdot \mathbf{y}) > \min\{f(\mathbf{x}), f(\mathbf{y})\} \tag{10}$$

All concave functions are quasiconcave, but not vice versa.
An alternative definition useful for economists:
For all $a \in R^1$, the set $\{\mathbf{x} : f(\mathbf{x}) \geq a\}$ is a "convex set".

- A set U is convex if $\forall \mathbf{x}, \mathbf{y} \in U$ and $t \in [0, 1]$, $t \cdot \mathbf{x} + (1 - t) \cdot \mathbf{y} \in U$.

Quasiconcavity is a deeper concept than concavity because it is preserved by monotonic transformations.
E.g. $f(x) = x^{1/2}$

- $f(x)$ is concave and quasiconcave
- But $f(x)^4 = x^2$
  - Is no longer concave
  - Remains quasiconcave

So quasiconcave is an "ordinal" feature of a function, unlike concavity.

Quasiconcavity and Optimization

- We will not show this, but quasiconcavity is the minimal assumption that ensures that a critical point is the global max of a differentiable function
- So if you know your function is quasiconcave, you do not need to check SOC: it will be satisfied.
- Since quasiconcavity is less strict than concavity **and** is an ordinal property, it is common for economists to assume objective functions are quasiconcave *when doing proofs*.
  - You will probably see that a lot in micro theory courses.
- In practice when *solving problems* with explicit utility functions, concavity is far more transparent in terms of derivatives, so that's usually focused on