Econ 205 - Slides from Lecture 13

Joel Sobel

September 10, 2010

Second Order Conditions

For a critical point $\mathbf{x}^*$ we have by Taylor's Theorem that

$$\frac{f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*)}{\|\mathbf{h}\|^2} = \frac{\frac{1}{2}\mathbf{h}^t D^2 f(\mathbf{z})\mathbf{h}}{\|\mathbf{h}\|^2}$$

We could write $\mathbf{h} = t\mathbf{v}$, where $\|\mathbf{v}\| = 1$. So

$$\begin{aligned}\frac{f(\mathbf{x}^* + t\mathbf{v}) - f(\mathbf{x}^*)}{t^2} &= \frac{\frac{1}{2}t^2\mathbf{v}^t D^2 f(\mathbf{z})\mathbf{v}}{t^2} \\ &= \frac{1}{2}\mathbf{v}^t D^2 f(\mathbf{z})\mathbf{v}\end{aligned}$$

for $\mathbf{z} \in B_{\|\mathbf{h}\|}(x^*)$

Sufficient Conditions

If f is twice continuously differentiable and $\mathbf{x}^*$ is a critical point of f, then:

1. $\mathbf{x}^*$is a local minimizer whenever the quadratic form $\mathbf{v}^t D^2 f(\mathbf{x}^*)\mathbf{v}$ is a positive definite.
2. $\mathbf{x}^*$is a local maximizer whenever the quadratic form $\mathbf{v}^t D^2 f(\mathbf{x}^*)\mathbf{v}$ is negative definite.

Necessary Conditions

If $D^2 f(\mathbf{x}^*)$ exists, then:

1. if $\mathbf{x}^*$ is a local maximizer, then $\mathbf{v}^t D^2 f(\mathbf{x}^*)\mathbf{v}$ is a negative semi-definite quadratic form.
2. if $\mathbf{x}^*$ is a local minimizer, then $\mathbf{v}^t D^2 f(\mathbf{x}^*)\mathbf{v}$ is a positive semi-definite quadratic form.

And indefinite?

Critical point cannot be either a maximum or a minimum.

Concavity

If f is concave, then $D^2 f(\mathbf{x})$ is negative semi-definite for all $\mathbf{x}$.

Example

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be

$$f(x, y) = 4x^3 + y^2 - 6xy + 6x$$

If (x^*, y^*) is a critical point, then

$$\nabla f(x^*, y^*) = \mathbf{0}$$

$$\frac{\partial f}{\partial x}(x, y) = 12x^2 - 6y + 6 = 0$$

$$\frac{\partial f}{\partial y}(x, y) = 2y - 6x = 0$$

from the second equation we get $2y = 6x$. Hence

$$\begin{aligned} & 12x^2 - 18x + 6 = 0 \\ \Rightarrow \quad & 2x^2 - 6x + 1 = 0 \\ \Rightarrow \quad & (2x - 1)(x - 1) = 0 \end{aligned}$$

the solutions are

$$x = \tfrac{1}{2} \quad \text{or} \quad x = 1$$

So we have 2 critical points

$$(\tfrac{1}{2}, \tfrac{3}{2}), (1, 3)$$

$$D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial x^2} \end{pmatrix}$$

$$= \begin{pmatrix} 24x & -6 \\ -6 & 2 \end{pmatrix}$$

So

$$D^2 f(\tfrac{1}{2}, \tfrac{3}{2}) = \begin{pmatrix} 12 & -6 \\ -6 & 2 \end{pmatrix}$$

This gives rise to an indefinite quadratic form (upper left entry positive, determinant negative). Hence $(\frac{1}{2}, \frac{3}{2})$ is neither a maximizer nor a minimizer.

$$D^2 f(1, 3) = \begin{pmatrix} 24 & -6 \\ -6 & 2 \end{pmatrix}$$

Since both leading principal minors are positive the point $(1, 3)$ must be a local minimum.

Convexity

Let $X \subset \mathbb{R}^n$.

Definition

x is an *interior point* of X if for some $\varepsilon > 0$

$$B_\varepsilon(x) \subset X$$

Definition

x is a *boundary point* of X if $\forall\, \varepsilon > 0$ both

$$B_\varepsilon(x) \cap X \neq \phi$$

and

$$B_\varepsilon(x) \cap [\mathbb{R}^n \backslash X] \neq \phi$$

Definition

X is called *open* if every $x \in X$ is an interior point.

Definition

X is called *closed* if it contains all of its limit points

Reminder: A hyperplane in $\mathbb{R}^n$ is of dimension $\mathbb{R}^{n-1}$ and can be expressed as

$$\mathbf{x} \cdot \mathbf{p} = c$$

where
$\mathbf{x} \in \mathbb{R}^n$
$\mathbf{p} \neq \mathbf{0}$ is in $\mathbb{R}^n$
$c \in \mathbb{R}$. $\mathbf{p}$ is the normal to the hyperplane.

Convex Sets

Definition

$X \subset \mathbb{R}^n$ is called *convex* if $\forall\, \mathbf{x}, \mathbf{y} \in X$ and $\forall\, \alpha \in [0,1]$ we have

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in X$$

Separating Hyperplane Theorem

Theorem

Given a nonempty, closed, convex set $X \subset \mathbb{R}^n$ *and* $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \notin X$*. There exists* $\mathbf{p} \in \mathbb{R}^n, \mathbf{p} \neq \mathbf{0}$*, and* $c \in \mathbb{R}$ *such that*

$$X \subset \{\mathbf{y} \mid \mathbf{y} \cdot \mathbf{p} \geq c\}$$

and

$$\mathbf{x} \cdot \mathbf{p} < c$$

That is, $\mathbf{y} \cdot \mathbf{p} = c$ defines a separating hyperplane for X.

1. Consider the problem of minimizing the distance between $\mathbf{x}$ and the set X.
2. Find $\mathbf{y}^*$ to solve:

$$\min \|\mathbf{x} - \mathbf{y}\| \text{ subject to } y \in X. \tag{1}$$

3. The norm is a continuous function.
4. While X is not necessarily bounded, since it is nonempty, we can find some element $\mathbf{z} \in X$. Without loss of generality we can replace X in (1) by $\{y \in X : \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{x}\|\}$. This set is compact because X is closed.
5. Hence there is a solution $\mathbf{y}^*$ to (1).

6. Let $\mathbf{p} = \mathbf{y}^* - \mathbf{x}$.
7. Since $\mathbf{x} \notin X$, $\mathbf{p} \neq \mathbf{0}$.
8. Let $c = \mathbf{y}^* \cdot \mathbf{p}$.
9. Since $c - \mathbf{p} \cdot \mathbf{x} = \|\mathrm{p}\|^2$, $c > \mathbf{p} \cdot \mathbf{x}$.

10. Suffices to show if

$$\mathbf{y} \in X, \text{ then } \mathbf{y} \cdot p \geq c.$$

11. This inequality is equivalent to

$$(\mathbf{y} - \mathbf{y}^*) \cdot (\mathbf{y}^* - \mathbf{x}) \geq 0. \tag{2}$$

12. Since X is convex and $\mathbf{y}^*$ is defined to solve (1), it must be that

$$\|t\mathbf{y} + (1-t)\mathbf{y}^* - x\|^2$$

is minimized when $t = 0$.

13. The derivative of $\|t\mathbf{y} + (1-t)\mathbf{y}^* - x\|^2$ is non-negative at $t = 0$.

14. Differentiation and simplifying yields inequality (2).

Extensions

- Without loss of generality you can normalize the normal to the separating hyperplane. That is, you can assume that $\|\mathbf{p}\| = 1$.
- If $\mathbf{x}$ is an element of the boundary of X, then you can approximate $\mathbf{x}$ by a sequence $\mathbf{x}_k$ such that each $\mathbf{x}_k \notin X$.
- This yields a sequence of $\mathbf{p}_k$, which can be taken to be unit vectors, that satisfy the conclusion of the theorem.
- A subsequence of the $\mathbf{p}_k$ must converge.
- The limit point $\mathbf{p}^*$ will satisfy the conclusion of the theorem (except we can only guarantee that $c \geq \mathbf{p}^* \cdot \mathbf{x}$ rather than the strict equality above).
- The closure of any convex set is convex.
- Given a convex set X and a point $\mathbf{x}$ not in the interior of the set, we can separate $\mathbf{x}$ from the closure of X.

Another Separating Hyperplane Theorem

Theorem

Given a convex set $X \subset \mathbb{R}^n$ *and* $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}$. *If* $\mathbf{x}$ *is not in the interior of* X, *then there exists* $\mathbf{p} \in \mathbb{R}^n, \mathbf{p} \neq \mathbf{0}$, *and* $c \in \mathbb{R}$ *such that*

$$X \subset \{\mathbf{y} \mid \mathbf{y} \cdot \mathbf{p} \geq c\}$$

and

$$\mathbf{x} \cdot \mathbf{p} \leq c$$

In the typical economic application, the separating hyperplane's normal has the interpretation of prices and the separation property states that a particular vector costs more than vectors in a consumption set.

Quasi-Concave and Quasi-Convex Functions

Definition

A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *quasiconcave* if $\{\mathbf{x} \mid f(\mathbf{x}) \geq a\}$ is convex for all $a \in \mathbb{R}$. In other words the upper contour set is a convex set.

Example

$f(\mathbf{x}) = -x_1^2 - x_2^2$
the upper contour set is WHAT?

Quasiconvex

Definition

A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *quasiconvex* if $\{\mathbf{x} \mid f(\mathbf{x}) \leq a\}$ is convex for all $a \in \mathbb{R}$. In other words the lower contour set is a convex set.

In the one-variable setting, we said that a function was quasiconcave if

$$\text{for all } x, y, \text{ and } \lambda \in (0,1), f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}. \tag{3}$$

If $a = \min\{f(x), f(y)\}$, then general definition implies (3). Conversely, if the condition in the definition fails, then there exists a, x, y such that $f(x), f(y) \geq a$ but $f(\lambda x + (1-\lambda)y) < a$. Condition (3) fails for these values of λ, x, and y.

Quasiconcavity and Quasiconvexity are global properties of a function. Unlike continuity, differentiability, concavity and convexity (of functions), they cannot be defined at a point.

Global Properties

Theorem

If f is concave and $\mathbf{x}^*$ is a local maximizer of f, then $\mathbf{x}^*$ is a global maximizer.

Proof.

Suppose $\mathbf{x}^*$ is not a global maximizer of f. Then there is a point $\widehat{\mathbf{x}}$ such that

$$f(\widehat{\mathbf{x}}) > f(\mathbf{x}^*)$$

but then

$$f(\alpha \mathbf{x}^* + (1-\alpha)\widehat{\mathbf{x}}) \geq \alpha f(\mathbf{x}^*) + (1-\alpha) f(\widehat{\mathbf{x}})$$

And this contradicts that $\mathbf{x}^*$ is a local max. □

Identifying Quasiconcave Functions

1. First solve for the level sets and graph a selection of them
2. Decide by inspection which side of the plotted level set is the upper contour set and which side is the lower contour set
3. Are the inspected upper and lower contour sets convex?

Relationship between Concavity and Quasiconcavity

Suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is concave $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then $\forall\, \alpha \in [0, 1]$

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \geq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

Examine the set $\{\, \mathbf{z} \mid f(\mathbf{z}) \geq c \}$.
Consider the points $\mathbf{x}, \mathbf{y} \in \{\, \mathbf{z} \mid f(\mathbf{z}) \geq c \}$,

$$\Longrightarrow f(\mathbf{x}) \geq c \quad \text{and} \quad f(\mathbf{y}) \geq c$$

and this means $\forall\, \alpha \in [0, 1]$ that

$$\begin{aligned} f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) &\geq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \\ &\geq \alpha c + (1-\alpha) c \\ &= c \end{aligned}$$

$$\Longrightarrow \alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in \{\, \mathbf{z} \mid f(\mathbf{z}) \geq c \}$$

Theorem

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is concave, then f is quasiconcave.
Similarly if f is convex, then f is quasiconvex.

e^x is quasiconcave but not concave. In fact it is also convex and quasiconvex.
because:

Theorem

Any increasing or decreasing function is both quasiconvex and quasiconcave.

Ordinal

Loosely: A concept will be thought of as *ordinal* if "order" is all that matters. For example the statement

$$f(\mathbf{x}) > f(\mathbf{y})$$

Definition

We say that the property of functions $*$ is ordinal if $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ has property $(*)$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is strictly increasing

$$\Longrightarrow g \circ f \text{ has property}(*)$$

In other words it doesn't matter how much greater the value $f(\mathbf{x})$ is than $f(\mathbf{y})$, it just matters that it is greater.

Cardinal

Definition

We say that the property of functions $*$ is cardinal if $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ has property $(*)$ and $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ is strictly increasing affine function, then

$$\Longrightarrow g \circ f \text{ has property}(*)$$

Concavity isn't ordinal

$$f : (0, \infty) \longrightarrow \mathbb{R}$$

$$f(x) = x^{\frac{1}{2}}$$

So if

$$x > y \Longrightarrow f(x) > f(y)$$

This is a concave function. Now suppose

$$g(y) = y^4$$

$$\Longrightarrow g \circ f(x) = x^2$$

And again we have that

$$x > y \Longrightarrow g \circ f(x) > g \circ f(y)$$

This is it now a convex function.

Quasiconcavity is Ordinal

Suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconcave. That is, the set $\{ x \mid f(x) \geq c \}$ is convex $\forall c \in \mathbb{R}$.
Now take a strictly increasing function

$$g : \mathbb{R} \longrightarrow \mathbb{R}$$

So the following 2 sets must be equivalent

$$\{ x \mid g \circ f(x) \geq c \} \iff \{ x \mid f(x) \geq g^{-1}(c) \}$$

and so the upper contour set remains convex $\implies$ quasiconcavity is preserved.

Insight

- Unconstrained optimization is easy because at an interior optima it must be that directional derivatives in all directions are zero.
- If you are at a maximum, then moving "forward" in any direction must not increase the objective function.
- For this section

$$D_v f(x) = \lim_{h \to 0^+} \frac{f(x + hv) - f(x)}{t}.$$

- So a one-sided directional derivative must be less than or equal to zero.
- Moving "backward" in that direction must also lower the function's value, so the one-sided directional derivative must be equal to zero.
- The same insight applies in the constrained case, but
- movement in some directions might be ruled out by constraints, FOCs are a set of inequalities that hold for all directions that are not ruled out by constraints.

Equality Constraints

The basic problem is:

$$\max f(x) \tag{4}$$

$$x \in S \tag{5}$$

- $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.
- The function f in (4) is called the objective function.
- We study maximization problems, but since minimizing a function h is the same as maximizing $-h$, the theory covers minimization problems as well.
- (5) is the constraint set.
- A solution to the problem is x^* such that (a) $x^* \in S$ and (b) if $x \in S$, then $f(x^*) \geq f(x)$.

- Calculus does not provide conditions for the existence of an optima (real analysis does, because continuous real-valued functions on compact sets attain their maxima).
- Calculus characterizes optima in terms of equations.
- When S is an open set, any maxima must be interior. Hence $Df = 0$.
- With constraints, this condition must be modified.

- Equality constraints:
  There exist $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that $S = \{x : g_i(x) = b_i\}$ for all i.
- Linear constraint set: $g_i(x) = a_i \cdot x$. In this case, it is important to have arbitrary constants on the right-hand side of the constraint.
- In linear case, b_i arbitrary.
- In general case, no lose of generality assuming $b_i = 0$, although it is interesting to ask how solutions change as RHS changes.

One Constraint

1. Use the constraint to write one variable in terms of the others.
2. Substitute out for that variable and solve the unconstrained problem.

Will this work?
Point (1) requires condition of IFT.

$$g(x_1, x_2) = c$$

implicitly define $x_2 = h(x_1)$ around x_1^* (the solution).
Recall that h is well defined if

$$\frac{\partial g}{\partial x_2}(\mathbf{x}^*) \neq 0$$

$$g(x_1, h(x_1)) = 0$$

$$\frac{\partial g}{\partial x_1}(\mathbf{x}^*) + \frac{\partial g}{\partial x_2}(\mathbf{x}^*) h'(x_1^*) = 0$$

$$h'(x_1) = -\frac{\frac{\partial g}{\partial x_1}(\mathbf{x}^*)}{\frac{\partial g}{\partial x_2}(\mathbf{x}^*)}$$

This gives us an unconstrained maximization problem, for $\mathbf{x}$ near $\mathbf{x}^*$.

$$\max_{x_1} f(x_1, h(x_1))$$

Because $\mathbf{x}^*$ solves the original maximization problem, it solves this one.

More generally . . .

Write $G = (g_1, \ldots, g_m)$ so that there are m constraints.
$G : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.
Typically, the problem is uninteresting if $m \geq n$ (because then the constraint set is likely to contain only one point or be empty).
The constraints are non-degenerate (or satisfy a regularity condition or a constraint qualification) at a point x^* if $DG(x^*)$ has full rank ($\{\nabla g_i(x^*)\}$ is a collection of m linearly independent vectors).

1. Divide the variable $x^* = (u^*, w^*)$ where $u^* \in \mathbb{R}^{n-m}$ and $w^* \in \mathbb{R}^m$ and $D_w G(x^*) \neq \mathbf{0}$.
2. The nondegeneracy condition guarantees that there are m variables you can select so that $D_w G(x^*)$ is nonsingular.
3. IFT implies it is possible to solve the equations $G(x^*) = 0$ for a function $W : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}^m$ such that $W(u^*) = w^*$ and $G(u^*, W(u^*)) = \mathbf{0}$ in a neighborhood of (u^*, w^*).
4. The constrained optimization problem is equivalent to solving the unconstrained optimization problem:

$$\max f(u, W(u)).$$

1. If x^* solves the original problem, then u^* must be a local maximum of the unconstrained problem.
2. Hence the composite function $h(u) = f(u, W(u))$ has a critical point at u^* (at a local max).
3. Chain rule: $Dh(u^*) = D_u f(x^*) + D_w f(x^*) DW(u^*)$.
4. IFT gives formula for $DW(u^*)$:

$$0 = Dh(u^*) = D_u f(x^*) - D_w f(x^*) \left(D_w G(x^*)\right)^{-1} D_u G(x^*) \tag{6}$$

or

$$D_u f(x^*) = D_w f(x^*) \left(D_w G(x^*)\right)^{-1} D_u G(x^*). \tag{7}$$

Let

$$\lambda = D_w f(x^*) \left(D_w G(x^*)\right)^{-1} \text{ or } D_w f(x^*) = \lambda D_w G(x^*).$$

It follows that from equation (7) that

$$D_u f(x^*) = \lambda D_u G(x^*),$$

which combined with $D_w f(x^*) = \lambda D_w G(x^*)$ we can write

$$Df(x^*) = \lambda DG(x^*).$$