ARE211, Fall2015

CALCULUS4: THU, SEP 17, 2015 PRINTED: SEPTEMBER 22, 2015 (LEC# 7)

Contents

2. Univariate and Multivariate Differentiation (cont)

Key points: Assume $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable

(1) If Hf is negative definite at $\mathbf{x}$, then f attains a strict local maximum at $\mathbf{x}$ iff $\bigtriangledown f(\mathbf{x}) = 0$ (see p. 4).

(2) If we replace "negative definite at $\mathbf{x}$" with "globally negative definite" then we can replace "local maximum" with "global maximum" in (1).

(3) global negative (semi) definiteness buys you a (weak) global max; local negative definiteness buys you nothing

(4) If f is concave and $\mathbb{C}^2$ then $\mathbf{Hf}$ is globally negative semi-definite

(5) If Hf is negative definite at $\mathbf{x}$, then the tangent plane to f at $\mathbf{x}$ is locally above the graph of f, and strictly above except at $\mathbf{x}$ (see p. 18).

(6) If we replace "negative definite at $\mathbf{x}$" with "globally negative definite" then we can replace "is locally" with "is globally" in (5).

(7) A sufficient condition for f to be quasi-concave is that at each $\mathbf{x}$ $\mathrm{H}\mathbf{f}(\mathbf{x})$ is globally negative definite on the subspace that is orthogonal to $\triangledown f(\mathbf{x})$.

In this lecture, we'll see the critical role that *continuous* differentiability plays in a number of important theorems. Based on what I said in earlier lectures, you might have been tempted to come to the conclusion that the only reason why economists insist on continuous differentiability is that this is a sufficient condition for *differentiability*, i.e., the vector space property of directional derivatives that I talked about. If this were the *only* reason why we cared about continuous differentiability, then it would be overkill. But in this lecture, we'll see that none of the theorems we care most about would be true if we had only imposed the weaker condition of differentiability.

2.4. **Taylor's Theorem (cont)**

Taylor's Theorem (continued): Why is the theorem so tremendously important? Because if you are only interested in the *sign* of $(f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}))$ and you have an n'th order Taylor expansion, then you know that for some neighborhood about $\bar{\mathbf{x}}$, the sign of your expansion will be the same as the sign of the true difference, i.e., $(f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}))$.

2.5. **Applying Taylor theory: 2nd order conditions for an unconstrained local max.**

Definition: A function $f : X \to \mathbb{R}$ attains a strict (weak) local maximum at $\mathbf{x} \in X$ if there exists $\epsilon > 0$ such that for all $\mathbf{y} \in X \cap B(\mathbf{x}, \epsilon)$, $f(\mathbf{y}) - f(\bar{\mathbf{x}}) < (\leq) \ 0$.

Going to be talking about necessary and sufficient conditions for a local optimum of a differentiable function.

Terminology is that first order conditions are *necessary* while second order conditions are *sufficient*.

The terms necessary and sufficient conditions have a formal meaning:

- If an event A cannot happen unless an event B happens, then B is said to be a *necessary condition* for A.
- If an event B *implies* that an event A will happen, then B is said to be a *sufficient condition* for A.

For example, consider a differentiable function from $\mathbb{R}^1$ to $\mathbb{R}^1$.

- f cannot attain an interior maximum at $\bar{x}$ *unless* $f'(\bar{x}) = 0$.
  - i.e., the maxmimum is A; the derivative condition is B.
  - Thus, the condition that the first derivative is zero is *necessary* for an interior maximum; called the first order conditions.

  - Emphasize strongly that this necessity business is delicate: derivative condition is only necessary provided that f is differentiable *and* we're talking interior maximum
- $f'(\bar{x}) = 0$ certainly doesn't IMPLY that f attains an interior maximum at $\bar{x}$
- If $f''(\bar{x}) < 0$, then the condition $f'(\bar{x}) = 0$ is both necessary and sufficient for an interior *local* maximum;
- Alternatively, if you know in advance that f is *strictly concave*, then the condition that $f'(\bar{x})$ is zero is necessary and sufficient for a *strict global* maximum.

Generalizing to functions defined on $\mathbb{R}^n$, a simple application of Taylor's theorem proves that if an appropriate local second order condition is satisfied, then the first order conditions are in fact necessary and sufficient for a *strict local* maximum

Theorem: Let X be an open subset of $\mathbb{R}^n$ and consider a twice continuously differentiable function $f : X \to \mathbb{R}$ and a point $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that the Hessian of f is negative definite at $\bar{\mathbf{x}}$. f attains a strict local maximum at $\bar{\mathbf{x}} \in \mathbb{R}^n$ iff $\bigtriangledown f(\bar{\mathbf{x}}) = 0$.

Proof of Necessity: To prove this part, we use the *local* version of Taylor's theorem. Suppose $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$. We'll show that in this case, $f(\cdot)$ cannot be maximized at $\bar{\mathbf{x}}$. Consider the first order Taylor expansion of f about $\bar{\mathbf{x}}$:

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) = \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \text{ a remainder term }. \tag{1}$$

Since X is open, there exists $\epsilon_1 > 0$ such that $B(\mathbf{x}, \epsilon_1) \subset X$. From the Taylor Young (Local Taylor) theorem there exists $\epsilon_2 \in (0, \epsilon_1)$ such that if $||\mathbf{dx}|| < \epsilon_2$, then the absolute value of the first term of the Taylor expansion is larger than the absolute value of the remainder term. Now, let $\mathbf{dx} = \lambda \bigtriangledown f(\bar{\mathbf{x}})$, where $\lambda > 0$ is picked sufficiently small that $||\lambda \bigtriangledown f(\bar{\mathbf{x}})|| < \epsilon_2$. Since the angle

θ between $\bigtriangledown f(\bar{\mathbf{x}})$ and $\mathbf{dx}$ is zero, we have $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} = \cos(\theta)|| \bigtriangledown f(\bar{\mathbf{x}})|| ||\mathbf{dx}|| > 0$. Therefore, $f(\bar{\mathbf{x}} + \mathbf{dx}) > f(\bar{\mathbf{x}})$. Since $\bar{\mathbf{x}} + \mathbf{dx} \in B(\bar{\mathbf{x}}, \epsilon_2) \subset B(\bar{\mathbf{x}}, \epsilon_1) \subset X$, $\bar{\mathbf{x}}$ does not maximize f on X. □

Proof of joint sufficiency: To prove sufficiency, we use the *global* version of Taylor's theorem. We need to show that there exists $\epsilon > 0$ such that for all $\mathbf{dx} \in B(0, \epsilon)$, $f(\bar{\mathbf{x}} + \mathbf{dx}) < f(\bar{\mathbf{x}})$. Let S denote the unit sphere, i.e., $S = \{\mathbf{v} \in \mathbb{R}^n : ||\mathbf{v}|| = 1\}$. Define the function $\delta(\cdot)$ on S by for $\mathbf{v} \in S$, $\delta(\mathbf{v}) = \mathbf{v}'\text{Hf}(\bar{\mathbf{x}})\mathbf{v}$. Since $\text{Hf}(\bar{\mathbf{x}})$ is negative definite, $\delta(\cdot) < 0$. Since $\delta(\cdot)$ is continuous and S is compact, $\delta(\cdot)$ attains a maximum on S. Let $\underline{\delta} = 0.5 \max(\delta(\cdot)) < 0$. Since f is twice continuously differentiable, there exists $\epsilon > 0$, such that for all $\mathbf{x}' \in B(\bar{\mathbf{x}}, \epsilon)$ and all $\mathbf{v} \in S$, $\mathbf{v}'\text{Hf}(\mathbf{x}')\mathbf{v} < \underline{\delta} < 0$. Hence for an *arbitrary vector* $0 \neq \mathbf{dx} \in \mathbb{R}^n$, $\mathbf{dx}'\text{Hf}(\mathbf{x}')\mathbf{dx} < 0$. Now pick $\mathbf{dx} \in B(0, \epsilon)$. Clearly for all $\lambda \in [0, 1]$, $\lambda\mathbf{dx} \in B(0, \epsilon)$ also, so that $\text{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})$ is negative definite; in particular $\frac{1}{2}\mathbf{dx}'\text{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx} < 0$. It now follows from the Taylor Lagrange (global Taylor) theorem that for some $\lambda \in [0, 1]$,

$$\begin{aligned} f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) &= \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\text{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx} \qquad (2) \\ &< \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} = 0. \end{aligned}$$

This completes the proof that $f(\bar{\mathbf{x}} + \mathbf{dx}) < f(\bar{\mathbf{x}})$. □.

One has to be extremely careful about the wording of these necessary and sufficient conditions: The following statement is FALSE: A function $f : \mathbb{R}^n \to \mathbb{R}$ attains a strict local maximum at $\bar{\mathbf{x}} \in \mathbb{R}^n$ iff $\bigtriangledown f(\bar{\mathbf{x}}) = 0$ and Hf is negative definite at $\bar{\mathbf{x}}$. The "if" part of this statement is true, but the "only if" isn't: you could have a strict local max at $\bar{\mathbf{x}}$ without f being negative definite at $\bar{\mathbf{x}}$, e.g. $-x^4$ attains a global max at 0 but it isn't negative definite at 0.

The theorem above only gives sufficient conditions for a *local* maximum. On the other hand, if $\text{Hf}(\cdot)$ is *globally* negative definite, we obtain the following global result.

Theorem: Consider a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. If $\mathrm{Hf}(\cdot)$ is globally negative (semi-)definite, f attains a strict (weak) *global* maximum at $\bar{\mathbf{x}} \in \mathbb{R}^n$ iff $\bigtriangledown f(\bar{\mathbf{x}}) = 0$.

The difference between this and the previous result (ignoring the 'semi' in parenthesees) is that that we have replaced negative definiteness at a point with global negative definiteness. This allows us to consider large $\mathbf{dx}$'s instead of just small ones, and obtain conditions for a *global* rather than just a local maximum.

Proof: If $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$, the theorem above establishes that $\mathbf{x}$ cannot be a local maximum, and hence it certainly isn't a global maximum. So necessity follows from the preceding theorem. Now suppose that $\bigtriangledown f(\bar{\mathbf{x}}) = 0$. Rewriting expression (2) above, we have that for *any* $\mathbf{dx} \in \mathbb{R}^n$

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx}, \text{ for some } \lambda \in [0,1] \tag{2}$$

which, since by assumption $\bigtriangledown f(\bar{\mathbf{x}}) = 0$,

$$= \quad \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx}$$

Since $\mathrm{Hf}(\cdot)$ is globally negative (semi)-definite, it follows that $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) < (\leq) \, 0$, proving that $f(\cdot)$ is globally strictly (weakly) maximized at $\mathbf{x}$. □

2.6. **The Hessian, the tangent plane and the graph of f**

Another important application of Taylor's theorem is the following result, which is obtained by essentially duplicating the proof of sufficiency above.

Theorem: Given $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable and $\bar{\mathbf{x}} \in \mathbb{R}^n$, if Hf is positive (negative) definite at $\bar{\mathbf{x}}$ then there exists $\epsilon > 0$ such that the tangent plane to f at $\bar{\mathbf{x}}$ lies *below* (*above*) the graph of the function on the ϵ-ball around $\bar{\mathbf{x}}$.

Proof: We'll do the case when Hf is negative definite at $\bar{\mathbf{x}}$. Once again, for any $\mathbf{dx} \neq 0$,

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) = \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx}, \text{ for some } \lambda \in [0,1] \tag{3}$$

Now, since $\mathrm{Hf}(\bar{\mathbf{x}})$ is negative definite by assumption, and $\mathrm{Hf}(\cdot)$ is continuous, there exists $\epsilon > 0$ such that if $\mathbf{dx} \in B(0,\epsilon)$ and $\lambda \in [0,1]$, then $\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})$ will be negative definite also. Hence the second term in the expression above will be negative. Subtracting $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}$ from both sides, we obtain that for all $\mathbf{dx}$ with $||\mathbf{dx}|| < \underline{\epsilon}$,

$$\underbrace{f(\bar{\mathbf{x}} + \mathbf{dx})}_{\text{the height of } f \text{ at } \bar{\mathbf{x}} + \mathbf{dx}} - \underbrace{\big(f(\bar{\mathbf{x}}) + \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}\big)}_{\text{the height of the tangent plane at } \bar{\mathbf{x}} + \mathbf{dx}} = \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx} < 0 \tag{4}$$

□

Notice that the theorem above isn't necessarily true without restricting $\mathbf{dx}$ to lie in a neighborhood of zero. Think of a camel. Put a tangent plane against the smaller hump, and the whole camel isn't underneath the plane. On the other hand, if $\mathrm{Hf}(\cdot)$ is *globally* negative definite, we can prove a stronger result.

Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ if twice continuously differentiable and $\mathrm{Hf}(\cdot)$ is negative (semi) definite on its entire domain, then for all $\mathbf{x} \in \mathbb{R}^n$, the tangent plane to f at $\mathbf{x}$ lies everywhere (weakly above) above the graph of the function.

Proof: The proof is identical to the proof of the preceding result except that we omit the caveat about $\mathbf{dx}$ being small. Since $\mathrm{Hf}(\cdot)$ is *everywhere* negative (semi) definite, then we know that regardless of the size of $\mathbf{dx}$, the matrix $\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})$ is, in particular, negative (semi) definite, so that the term $\frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}} + \lambda\mathbf{dx})\mathbf{dx}$ in expression (2) above is negative (non-positive). Hence the left hand side of expression (4) above is also negative (non-positive). □

Finally, we prove that for twice-differentiable functions, concavity and global negative-semi-definiteness of the Hessian are equivalent conditions.

Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable then f is concave if and only if $\mathrm{Hf}(\cdot)$ is negative semi-definite on its entire domain.

Proof: The necessity part follows easily from the above theorem. I'll leave it as an exercise. Sufficiency is harder. We'll prove that if f is not concave then $Hf(\cdot)$ is not globally semi-definite. Suppose that f is not concave, i.e., that there exists $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\lambda' \in (0,1)$ such that $f(\lambda'\mathbf{y} + (1-\lambda')\mathbf{z}) < \lambda' f(\mathbf{y}) + (1-\lambda') f(\mathbf{z})$. For each $\lambda \in [0,1]$, let $g(\lambda) = f(\lambda\mathbf{y} + (1-\lambda)\mathbf{z}) - \big[\lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{z})\big]$. Note that $g(0) = g(1) = 0$, while by assumption $g(\lambda') < 0$. Moreover, since the term in square brackets is affine, it does not contribute to $g''(\lambda)$. Indeed, applying the chain rule.

$$g''(\lambda) \quad = \quad \frac{d^2}{d\lambda^2}\big(f(\lambda\mathbf{y} + (1-\lambda)\mathbf{z})\big) \quad = \quad (\mathbf{y}-\mathbf{z})' Hf(\lambda\mathbf{y} + (1-\lambda)\mathbf{z})(\mathbf{y}-\mathbf{z})$$

Since $[0,1]$ is compact and g is continuous, g attains a global minimum on $[0,1]$. Let $\bar{\lambda}$ be a minimizer of g on $[0,1]$. Since $g(\bar{\lambda}) \le g(\lambda') < 0$, $\bar{\lambda} \in (0,1)$. Hence $g'(\bar{\lambda}) = 0$. Let $d\lambda = -\bar{\lambda}$ so that $\bar{\lambda} + d\lambda = 0$. By Global Taylor, there exists $\lambda \in [0, \bar{\lambda}]$ such that

$$g(0) - g(\bar{\lambda}) = g'(\bar{\lambda})d\lambda + 0.5g''(\lambda)d\lambda^2 = 0.5g''(\lambda)d\lambda^2 = 0.5d\lambda^2(\mathbf{y}-\mathbf{z})' Hf(\lambda\mathbf{y} + (1-\lambda)\mathbf{z})(\mathbf{y}-\mathbf{z}) > 0$$

Hence we have established that $Hf(\cdot)$ is not negative semi-definite at $\lambda\mathbf{y} + (1-\lambda)\mathbf{z}$. □

2.7. Quasiconcavity, quasiconvexity: an informal review

This subsection and subsection 2.8 below are taken from very old notes, providing an informal graphical introduction to quasi-ness. They should be viewed as a digression from the main theme of these notes, included for review purposes.

Recall that convex and concave functions were characterized by whether or not the area above or below their graphs were convex. There's another class of functions that are characterized by whether or not their *upper or lower contour sets* are convex.

It turns out that for economists, the critical issue about a function is not whether it is convex or concave, but simply whether its *upper or lower contour sets* are convex.

This should be very familiar to you: recall that what matters about a person's utility function is not the *amount* of utility that a person receives from different bundles but the shape of the person's indifference curves. Well, indifference curves are just the level sets. Notice that you always draw functions that have convex upper contour sets: called the law of diminishing marginal rate of substitution.

So whether or not a function is concave or not turns out to be of relatively minor importance to economists. Consider Fig. 1. Though it's not entirely clear from the picture, the function graphed here has a striking resemblance to the concave function in the preceding graph: *the two functions have exactly the same level sets.* The second function is clearly not concave, but from an economic standpoint it works just as well as the concave function.

Definition: A function is quasiconcave if *all* of its upper contour sets are convex.

Definition: A function is quasiconvex if *all* of its lower contour sets are convex.

So in most of the economics you do, the assumption you will see is that utility functions are quasi-concave.

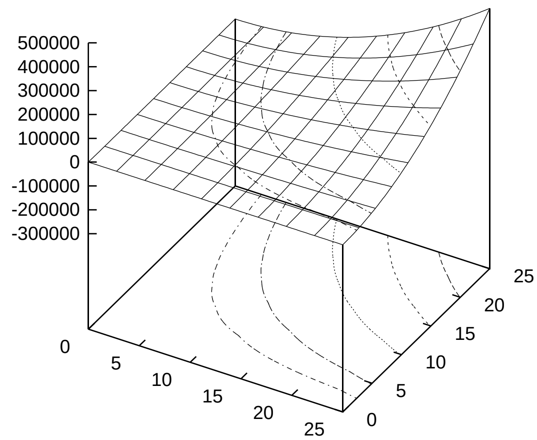

FIGURE 1. Level and contour sets of a quasiconcave function

FIGURE 2. A concave, quasi concave and a neither funtion

Most people find this concept rather difficult even though they are quite used to assuming diminishing mrs. A good test of whether you understand quasi concavity or not is to look at the concept in one dimension. Fig. 2 illustrates a concave, quasiconcave and a not quasiconcave function of one variable.

2.8. **Strict Quasiconcavity: an informal review.**

This subsection and subsection 2.7 above are taken from very old notes, providing an informal graphical introduction to quasi-ness. They should be viewed as a digression from the main theme of these notes, included for review purposes.

Just as there are strictly concave vs weakly concave functions or just concave functions, there are also strictly quasiconcave vs quasiconcave functions. The definition of strict quasi-concavity is less clean than the definition of quasi-concavity, but the properties of strictly quasi-concave functions are MUCH cleaner.

Definition: A function f is **strictly quasi-concave** if for any two points $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \neq \mathbf{y}$, in the domain of f, whenever $f(\mathbf{x}) \leq f(\mathbf{y})$, then f assigns a value strictly higher than $f(\mathbf{x})$ to every point on the line segment joining $\mathbf{x}$ and $\mathbf{y}$ except the points $\mathbf{x}$ and $\mathbf{y}$ themselves. Thus, s.q.c., rules out quasi-concave functions that have:

- straight line level sets (pyramids).
- flat spots (pyramids with helipads on the top).

Note that the above definition would not work if we replaced "$f(\mathbf{x}) \leq f(\mathbf{y})$" with "$f(\mathbf{x}) = f(\mathbf{y})$." More specifically, consider the following, similar but weaker condition: "for any two points $\mathbf{x}$ and $\mathbf{y}$ in the domain of f, whenever $f(\mathbf{x}) = f(\mathbf{y})$, then f assigns a value strictly higher than $f(\mathbf{x})$ to every point on the open line segment stricly between $\mathbf{x}$ and $\mathbf{y}$." This condition is satisfied by *any* function which has the property that no point in the range is reached from more than one point in the domain. E.g., consider the function defined on $\mathbb{R}$ by $f(x) = 1/x$, for $x \neq 0$; $f(0) = 0$. This function is not even quasi-concave, and so certaintly not strictly so: To see that it's not quasi-concave, note that the levelset corresponding to -1 is $[\infty, -1] \cup \mathbb{R}_+$, which is not a convex set. But since there are no points $\mathbf{x}$ and $\mathbf{y}$ s.t. $f(\mathbf{x}) = f(\mathbf{y})$, the function satisfies the latter condition trivially. (Thanks to Rob Letzler (2001) for this example.)

Definition: A function is strictly quasiconcave if all of its *upper* contour sets are strictly convex sets and *none* of its *level* sets have any width (i.e., no interior).

Definition: A function is strictly quasiconvex if all of its *lower* contour sets are strictly convex sets and *none* of its *level* sets have any width (i.e., no interior).

The first condition rules out straight-line level sets while the second rules out flat spots.

Two questions: Why do economists care so much about quasi-concavity? What is this long discussion doing in an overview of optimization theory?

The answer to both questions is that quasi-concavity is almost, but not quite, as good as concavity in terms of providing second order conditions for a maximum. Recall that if f is concave, then a necessary and sufficient condition for f to attain a global maximum at $\mathbf{x}^*$ is that the *first order conditions* for a max are satisfied at $\mathbf{x}^*$.

We can almost, but not quite, replace the word concavity by quasiconcavity in the above sentence. "Almost," however, is a very large word in mathematics: it's *not* true that if f is quasiconcave, then a necessary and sufficient condition for f to attain a global maximum at $\mathbf{x}^*$ is that the first order conditions for a max are satisfied at $\mathbf{x}^*$.

For example, consider the problem: max $f(x) = x^3$ on $[-1, 1]$, the function is strictly quasi-concave, and the first order conditions for a max are satisfied at $x = 0$, but it doesn't attain a max/min or anything at 0.

The following is true however: if f is quasiconcave *and the gradient of f never vanishes*, then a necessary and sufficient condition for f to attain a global maximum at $\mathbf{x}^*$ is that the first order conditions for a max are satisfied at $\mathbf{x}^*$.

Observe how the caveat takes care of the nasty example.

Now note that while quasi-concavity is a very useful second order condition for a *constrained* maximum, it's a true but pretty useless one for an *unconstrained* max.

The following statement is certainly true, but not particularly helpful. In fact it's particularly useless. If f is quasiconcave and the gradient of f never vanishes, then a necessary and sufficient condition for f to attain an *unconstrained* global maximum at $\mathbf{x}^*$ is that the first order conditions for a max are satisfied at $\mathbf{x}^*$.

Why isn't this particularly helpful?

Quasi-concavity can, however, provide some help in an unconstrained maximization problem, because it guarantees that a strict local maximum is a global maximum. That is, if you know your function is quasi-concave, then if your first order conditions and *local* second order conditions for a *strict* local maximum are satisfied, you know much more than you would know if you didn't know that your function were quasi-concave.

Fact: : If f is quasi-concave, a strict local maximum is a strict global maximum.

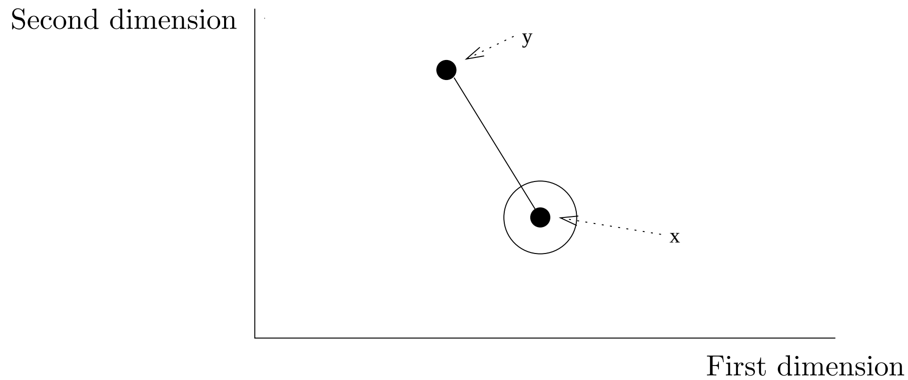

FIGURE 3. A function with a strict local maximum that is not a strict global maximum

Proof: : Consider a function with a strict local maximum that isn't a strict global maximum. We'll show that the function can't be quasiconcave. (Common way to prove things in math: showing "not B implies not A" is equivalent to (and often much easier than) showing that "A implies B". In this case "B" is the existence of a unique global maximum. "A" is the q.c.ness of the function. We're showing not B implies not A.

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ which has a strict local max at $\mathbf{x}$, but there exists a vector $\mathbf{y} \in \mathbb{R}^2$ such that f is at least as high at $\mathbf{y}$ as at $\mathbf{x}$.

- Consider the level set that goes through $\mathbf{x}$.
- As we've seen before, the strict local max $\mathbf{x}$ is an isolated point on this level set (though there may be other points further away that belong to the level set through $\mathbf{x}$).
- The point $\mathbf{y}$ must belong to the upper contour set corresponding to $f(\mathbf{x})$ (because f is at least as high at $\mathbf{y}$ as it is at $\mathbf{x}$.)
- Join up the line between $\mathbf{x}$ and $\mathbf{y}$.
- But we know that at least a part of this line doesn't belong to the upper contour set corresponding to $f(\mathbf{x})$, because f is larger at $\mathbf{x}$ than everywhere in a nbd of $\mathbf{x}$. Remember

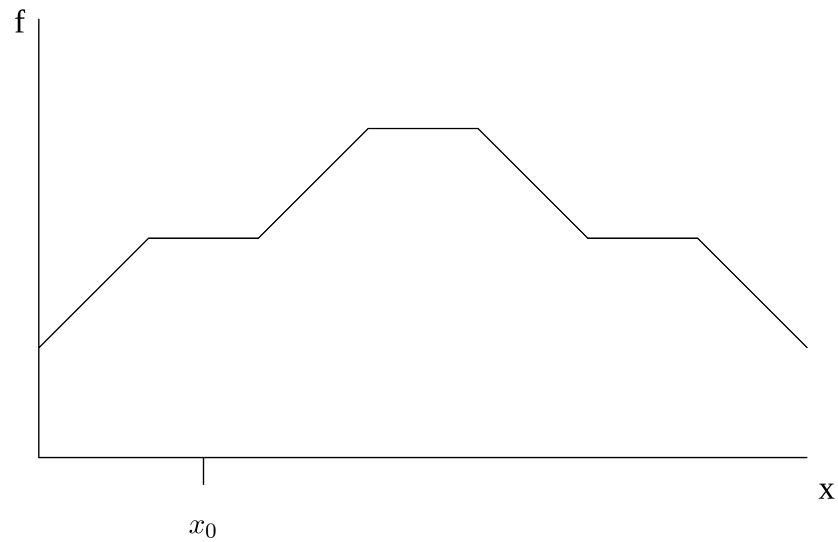

FIGURE 4. Quasi-concavity doesn't imply local max is a global max

the upper contour set lives in the domain of the function i.e. in the horizontal plane. Conclude that the upper contour set corresponding to $f(\mathbf{x})$ is *not* a convex set. Therefore, $f(\cdot)$ is not a quasiconcave function.

Note that it is *not true* that a weak local maximum of a quasi-concave function is necessarily a global max. Fig. 4 provides an example of a quasi-concave function f with lots of local maxima that are not global maxima. For example, f attains a local max at x_0.

Recapitulate: economists focus on quasi-concave functions because they have precisely the property that they care about:

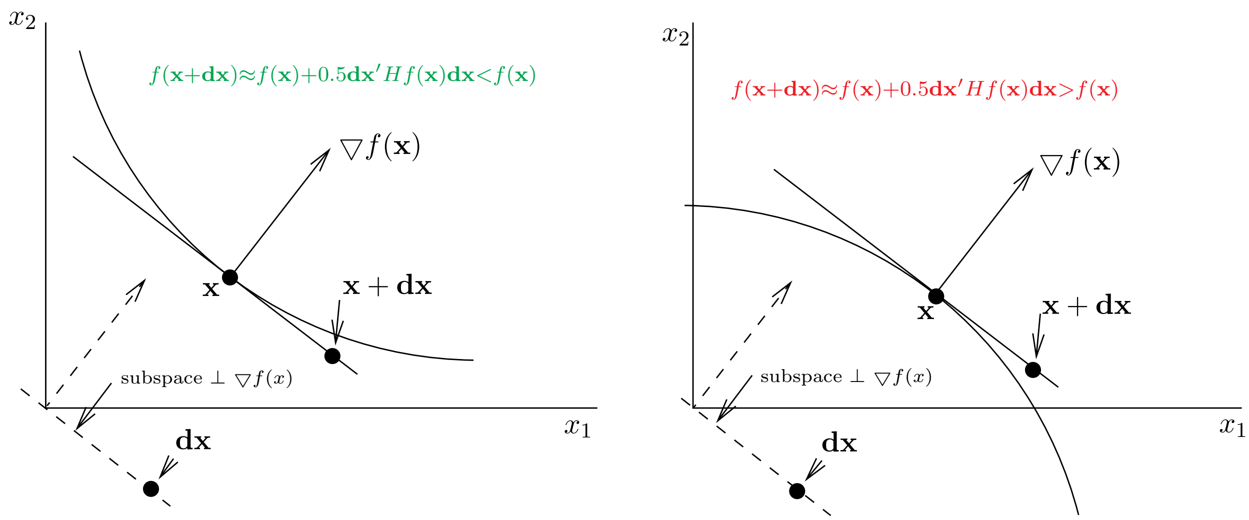

FIGURE 5. Negative and positive definiteness on subspace tangent to gradient

(a) first order necessary conditions for a *constrained* maximum of a quasi-concave function are not only necessary but also sufficient for a constrained max. provided you add the caveat that the gradient never vanishes. In economics, the label for this is *local non-satiation*.

(b) for any quasi-concave function, a strict local maximum is a strict global maximum.

2.9. **Sufficient conditions for quasi-concavity: intuitive review**

We will prove in this subsection that a *sufficient* condition for a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ to be quasi-concave is that for all $\mathbf{x}$, $\mathrm{Hf}(\mathbf{x})$ is negative semidefinite *on the subspace of* $\mathbb{R}^n$ *which is orthogonal to the gradient of* f. Formally the sufficient condition is:

$$\text{for all } \mathbf{x} \text{ and all } \mathbf{dx} \text{ such that } \bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}=0, \mathbf{dx}'\mathrm{Hf}(\mathbf{x})\mathbf{dx}\le 0. \qquad (5)$$

Later we'll see an equivalent representation of condition (5) in terms of the minors of *bordered Hessians*.

To prove that (5) is sufficient for quasi-concavity, we first need a lemma. Suppose we are maximizing or minimizing a function f on an *open* line segment L. A necessary condition for a point to be an extremum (i.e., maximum or minimum) on L is that the gradient of f at that point is orthogonal to the line segment. Thus result is really nothing more then a special case of the standard KKT conditions which we will learn in the next topic—i.e., maximizing an objective function on a line—but since the format of the result is slightly different from the one that we'll learn—when we learn the KKT conditions, the constraint set will be written as the intersection of lower contour sets of quasi-concave functions, here it's just a line—I'm going to state and prove it as a separate result.

Lemma: Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Fix $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and let $L = \{\lambda\mathbf{x} + (1-\lambda)\mathbf{y} : \lambda \in (0,1)\}$. A necessary condition for $\bar{\mathbf{x}} \in L$ to be an extremum on L is that $\bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} = 0$, for any $\mathbf{dx}$ such that $\bar{\mathbf{x}} + \mathbf{dx} \in L$.

Proof: We'll deal only with *maxima*; the argument for minima is parallel. Suppose $\bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} \neq 0$, for some $\mathbf{dx}$ such that $\bar{\mathbf{x}} + \mathbf{dx} \in L$. Assume w.l.o.g. that $\bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} > 0$ (if not, we could pick $-\mathbf{dx}$). By continuity, there exists a neighborhood U of $\bar{\mathbf{x}}$ such that for all $\mathbf{x}' \in U$, $\bigtriangledown f(\mathbf{x}') \cdot \mathbf{dx} > 0$. Pick $\psi > 0$ sufficiently small that $\bar{\mathbf{x}} + \psi\mathbf{dx} \in U$. By Global Taylor, there exists $\lambda \in [0,1]$ such that

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) = \bigtriangledown f(\bar{\mathbf{x}} + \lambda\psi\mathbf{dx})\psi\mathbf{dx}$$

But since $\bar{\mathbf{x}} + \psi\mathbf{dx}) \in U$, $\bar{\mathbf{x}} + \lambda\psi\mathbf{dx} \in U$ also, hence

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}} + \lambda\psi\mathbf{dx})\psi\mathbf{dx} \quad > \quad 0$$

Moreover, $\bar{\mathbf{x}} + \psi\mathbf{dx} \in L$ so $f(\cdot)$ is not maximized on L at $\bar{\mathbf{x}}$. □

We can now prove the relationship bewteen quasi-concavity and global semi-definiteness subject to constraint.

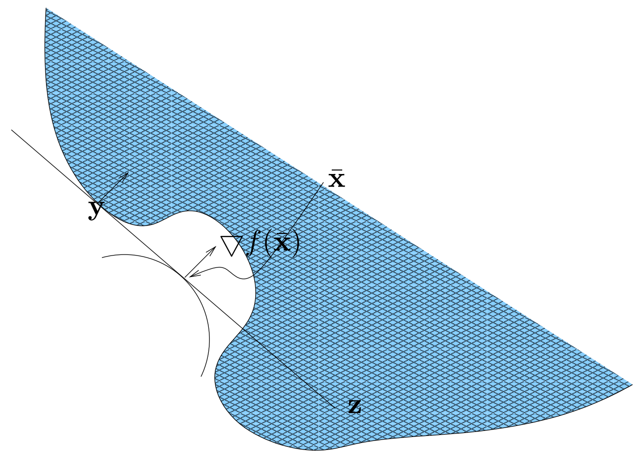

FIGURE 6. Negative definiteness subject to constraint implies QC

Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ is thrice continuously differentiable, then (5) is a sufficient condition for f to be quasi-concave.

Recall that by definition, f is quasi-concave if all upper contour sets are convex. We'll prove that if f is *not* quasi-concave then (5) is violated. Our approach will be to find a point $\mathbf{x}$ and a shift $\mathbf{dx}$ such that $\bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx} = 0$ but $\mathbf{dx}'\mathrm{Hf}(\mathbf{x})\mathbf{dx} > 0$, violating (5). Fig. 6 shows how to find $\mathbf{x}$: we find the *minimum* $\bar{\mathbf{x}}$ of f on the line segment joining $\mathbf{y}$ and $\mathbf{z}$, and then invoke the Lemma above, which says that the gradient of f at $\bar{\mathbf{x}}$ must be orthogonal to the line segment. Then, we can use Taylor to establish that the level set corresponding to $f(\bar{\mathbf{x}})$ *must, locally,* look something like the dashed line, and we're done. The above isn't quite precise, but the proof below is.

Proof: Assume that f is not quasi-concave, i.e., one of its upper contour sets is not convex. In this case, there must exist $\mathbf{y}, \mathbf{z}, \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} = \psi\mathbf{y} + (1-\psi)\mathbf{z}$, for some $\psi \in (0,1)$, and $f(\mathbf{x}) < f(\mathbf{y}) \leq f(\mathbf{z})$. We will now show that property (5) is not satisfied. Let $\bar{L} = \{\psi\mathbf{y} + (1-\psi)\mathbf{z} : \psi \in [0,1]\}$ and let L denote the interior of $\bar{L}$. (Note that $\bar{L}$ is a *closed* interval, cf the open interval L in the Lemma above.) Since f is continuous and $\bar{L}$ is compact, f attains a (global) minimum on $\bar{L}$. Clearly, since by assumption $f(\mathbf{x}) < \min[f(\mathbf{y}), f(\mathbf{z})]$, it follows that the global *minimizer* $\bar{\mathbf{x}}$ of f on $\bar{L}$ necessarily belongs also to L. There could, however, be multiple minimizers of f on L. In this case, pick a minimizer on the boundary of the minimal set. Since L is open, we can pick $\mathbf{dx} \neq 0$ such that $\bar{\mathbf{x}} + \mathbf{dx} \in L$, and by construction $f(\bar{\mathbf{x}} + \mathbf{dx}) > f(\bar{\mathbf{x}})$. By the Local Taylor theorem,

$$0 \quad < \quad f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} + \mathrm{Rem},$$

But from the Lemma above, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} = 0$ so that

$$0 \quad < \quad \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} + \mathrm{Rem},$$

If $||\mathbf{dx}||$ sufficiently small, $|\frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx}| > |\mathrm{Rem}|$, hence $\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} > 0$. But $\bigtriangledown f(\bar{\mathbf{x}})\lambda\mathbf{dx} = 0$, hence condition (5) is violated. We have thus established that (5) is sufficient for quasi-concavity.

□

2.10. **Terminology Review**

Keep these straight.

- The derivative of f at $\bar{\mathbf{x}}$: this is a point (which may be a scalar, vector or matrix).
  - partial derivative of f w.r.t variable i at $\bar{\mathbf{x}}$.
  - directional derivative of f in the direction h at $\bar{\mathbf{x}}$.
  - crosspartial derivative of f_i w.r.t variable j at $\bar{\mathbf{x}}$.
  - total derivative of f w.r.t variable i at $\bar{\mathbf{x}}$. The total derivative is different from the partial derivative w.r.t. i iff other variables change as the i'th variable changes. In this case, the change in the other variables determine a *direction*; e.g., if q depends on p, then a change in p induces a change in (p, q)-space in the direction $h = (1, q'(p))$. However, whenever $q'(p) \neq 0$, the total derivative is *strictly larger* than the directional derivative in the direction h.

- The derivative of f: this is a function. It's the generic term
  - functions from $\mathbb{R}^1$ to $\mathbb{R}^1$: derivative is usually just called the derivative.
  - functions from $\mathbb{R}^n$ to $\mathbb{R}^1$: derivative is usually called the gradient.
  - functions from $\mathbb{R}^n$ to $\mathbb{R}^m$: derivative is called the Jacobian matrix
  - the Hessian of f is the Jacobian of the derivative of f: i.e., the matrix of 2nd partial derivatives.
- The differential of f at $\bar{x}$: this is also a function, but a different one from the gradient. It's the unique linear function whose coefficient vector is the gradient vector evaluated at $\bar{x}$. Again, the differential may be a linear mapping from $\mathbb{R}^1$ to $\mathbb{R}^1$, from $\mathbb{R}^n$ to $\mathbb{R}^1$, or from $\mathbb{R}^n$ to $\mathbb{R}^m$, depending on the domain and codomain of the original function f.