2. Graphical Overview of Optimization Theory (cont)

2.6. Level Sets, upper and lower contour sets and Gradient vectors (cont)

Vectors: Recall that a vector in $\mathbb{R}^n$ is a collection of $n$ scalars. A vector in $\mathbb{R}^2$ is often depicted as an arrow. Properly the base of the arrow should be at the origin, but often you see vectors that have been “picked up” and placed elsewhere. Example below.

Gradient vectors: When economists draw level sets through a point, they frequently attach arrows to the level sets. These arrows are pictorial representation of the gradient vector, i.e., the slope of $f$ at $x$, $f'(x)$. Its components are the partial derivatives of the function $f$, evaluated at $x$, i.e., $(f_1(x), \cdots, f_n(x))$
Example: \( f(x) = 2x_1x_2 \), evaluated at \((2, 1)\), i.e., \( f'(2, 1) = (2x_2, 2x_1) = (2, 4) \). Draw the level set through \((2, 1)\), draw the gradient through the origin, lift it up and place its base at \((2, 1)\). Generally, the gradient of a function with \( n \) arguments is a point in \( \mathbb{R}^n \), and for this reason, you often see the gradient vector drawn in the domain of the function, e.g., for functions in \( \mathbb{R}^2 \), you often draw the gradient vector in the horizontal plane.

The gradient vector points in the direction of steepest ascent: Consider Fig. 18. Let \( x \) denote the point in the domain where the first straight line touches the circle. The graph represents a nice symmetric mountain which you are currently about to scale. You are currently at the point \( x \). You’re a macho kind of person and you want to go up the mountain in the steepest way possible. Ask yourself the question, looking at the figure. What direction from \( x \) is the steepest way up?

Answer is: the direction perpendicular to the straight line. Draw an arrow pointing in this direction. Now the gradient vector of \( f \) at \( x \) is an arrow pointing in precisely the direction you’ve drawn.

The following things about the gradient vector are useful to know:
• its length is a measure of the steepness of the function at that point (i.e., the steeper the function, the longer is the arrow.)

• as we’ve seen it is perpendicular to the level set at the point \( x \)

• it points inside the upper contour set. **Note Well: It could point into the upper contour set, but then go out the other side!**

• as we’ve seen, it points in the direction of steepest ascent of the function.

When we get to constrained optimization, we’ll talk a lot more about this vector.

2.7. **Quasiconcavity, quasiconvexity**

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. 19. Though it’s not entirely clear from the picture, the function graphed
Figure 19. Level and contour sets of a quasiconcave function

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.

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. 20 illustrates a concave, quasiconcave and a not quasiconcave function of one variable.
2.8. **Strict Quasiconcavity**

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 $x$ and $y$, $x \neq y$, in the domain of $f$, whenever $f(x) \leq f(y)$, then $f$ assigns a value strictly higher than $f(x)$ to every point on the line segment joining $x$ and $y$ except the points $x$ and $y$ 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 “$f(x) \leq f(y)$” with “$f(x) = f(y)$.” More specifically, consider the following, similar but weaker condition: “for any two points $x$ and
y in the domain of $f$, whenever $f(x) = f(y)$, then $f$ assigns a value strictly higher than $f(x)$ to every point on the open line segment strictly between $x$ and $y.$" 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 certainly 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 $x$ and $y$ s.t. $f(x) = f(y)$, 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 $x^*$ is that the *first order conditions* for a max are satisfied at $x^*$. 
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 $x^*$ is that the first order conditions for a max are satisfied at $x^*$.

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 $x^*$ is that the first order conditions for a max are satisfied at $x^*$.

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 $x^*$ is that the first order conditions for a max are satisfied at $x^*$.

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.

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 \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) which has a strict local max at \( x \), but there exists a vector \( y \in \mathbb{R}^2 \) such that \( f \) is at least as high at \( y \) as at \( x \).

- Consider the level set that goes through \( x \).
- As we’ve seen before, the strict local max \( x \) is an isolated point on this level set (though there may be other points further away that belong to the level set through \( x \)).
• The point $y$ must belong to the upper contour set corresponding to $f(x)$ (because $f$ is at least as high at $y$ as it is at $x$.)

• Join up the line between $x$ and $y$.

• But we know that at least a part of this line doesn’t belong to the upper contour set corresponding to $f(x)$, because $f$ is larger at $x$ than everywhere in a nbd of $x$. Remember 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 $f(x)$ 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. 22 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:

(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. Constrained Optimization: Several Variables

As I’ve said before, economists almost never solve unconstrained maximization problems. The whole mission of economics is to optimize subject to limited resources, and the economic limitations correspond to mathematical constraints.

Throughout this entire subsection, we’re going to assume that all the functions we have to deal with are as differentiable as we need them to be. As before, we’re going to study some “exceptions to the rule,” but unlike in the first graphical lecture, the source of these exceptions is not going to be non-differentiability.

There are two kinds of constraints: equality constraints and inequality constraints. We’ll consider a few examples.

(1) Maximizing subject to being on a line; Economics 101 solution to consumer’s utility maximization subject to a budget constraint.

(2) Maximizing subject to being on a line plus nonnegativity constraints. Consumer’s problem has the additional condition that you can’t consume negative quantities.

(3) Maximizing subject to being inside a convex set (partly defined by nonnegativity constraints.)

Theme of this section. All nonlinear programming problems (NPP) are essentially the same. Focus on the necessary condition for a constrained max: relates the arrows associated with the level sets of the objective function (gradient vector of objective function) with the arrows associated with the level sets of the constraint functions (gradient vectors) that are satisfied with equality. So there’s really only one thing you have to know in order to understand NPP. (You need to know quite a
lot more to actually go out and compute a solution to a real NPP, but if all you want to do is understand NPP theory, this is all you need to know.)

Talk about nonnegative and positive cones:

- Math cones are like icecream cones only longer: infinitely long in fact.
- Draw a couple of vectors: the cone consists of all the vectors that are inside the boundaries defined by these vectors.
- Indicate the nonnegative cone defined by them: the vectors that are inside the cone.
- Draw a vector inside the cone
- Ask what is the positive cone defined by a single vector.
- Note that the positive and the nonnegative cones defined by \( \{x, -x\}, x \neq 0 \), is the entire line through \( x \) (including zero). That is, in this case, the positive and nonnegative cones are identical.
- Draw three cones in the plane, observe that the inside one will be redundant, i.e., unnecessary.

While this section is supposed to not involve any symbols, it’s difficult to understand precisely what a cone is without them. So here are formal symbolic definitions of the various kinds of cones. (Recall that \( \mathbb{R}_+ \) denotes the set of nonnegative scalars and \( \mathbb{R}_{++} \) denotes the set of positive scalars, etc.)
Definition: Given a set of $m$ vectors, $V = \{v^1, ..., v^m\}$ in $\mathbb{R}^n$,

1. the nonnegative cone generated by $V$ is defined as
   
   \[ C^+(V) = \{ w \in \mathbb{R}^n : \exists t \in \mathbb{R}_+^m \text{ s.t. } w = \sum_{k=1}^m t_k v^k \}. \]

2. the positive cone generated by $V$ is defined as
   
   \[ C^{++}(V) = \{ w \in \mathbb{R}^n : \exists t \in \mathbb{R}_+^m \text{ s.t. } w = \sum_{k=1}^m t_k v^k \}. \]

3. the nonpositive cone generated by $V$ is defined as
   
   \[ C^-(V) = \{ w \in \mathbb{R}^n : \exists t \in \mathbb{R}_-^m \text{ s.t. } w = \sum_{k=1}^m t_k v^k \}. \]

4. the negative cone generated by $V$ is defined as
   
   \[ C^{--}(V) = \{ w \in \mathbb{R}^n : \exists t \in \mathbb{R}_-^m \text{ s.t. } w = \sum_{k=1}^m t_k v^k \}. \]

Mantra: (Except for a couple of bizarre exceptions) a necessary condition for $x^*$ to solve a constrained maximization problem is that the gradient vector of the objective function at $x^*$ belongs to the nonnegative cone generated by the gradient vector(s) of the constraint(s) that are satisfied with equality at $x^*$. (Parenthetical remark: Suppose that $x^*$ solves the maximization problem, and let $J$ denote the set of constraints that are satisfied with equality at $x^*$. If both the gradient of the objective, and the gradients of each constraint in $J$, are nonzero at $x^*$, then the gradient of the objective also belongs to the positive cone generated by the gradient vector(s) of the constraint(s) in the set $J$ that are binding.¹)

The word “binding” is a controversial word in the world of math for economics classes. By a “binding constraint” we mean a constraint with the following property: if you relax the constraint a little bit, then the maximized value of the objective function will increase. For example, consider

¹ To see why the caveat about the gradients being nonzero is required, consider the following two problems: (a) max $f(x) = x^3$ s.t. $x \leq 0$; (b) max $f(x) = x$ s.t. $x^3 \leq 0$. In both examples, the solution to the problem is at $x = 0$ and the constraint is binding. However, in (a), $f'(0) = 0$, while in (b), $f'(0) = 1$ but the gradient of the constraint is zero. Thus in neither case does the gradient of the objective belong to the positive cone generated by the gradient of the binding constraint.
the problem

$$\max_x u(x) = \sqrt{x_1 x_2} \quad \text{s.t.} \quad p \cdot x \leq y, x \geq 0.$$ 

The solution to this problem is: $$x_i = y/(2p_i), \ i = 1, 2.$$ In this example, we say that the budget constraint, $$p \cdot x \leq y,$$ is \textit{binding}. Intuitively, we mean by this that increasing income $$y$$ by a little bit will increase utility $$u$$. (By “a little bit,” we mean “by an arbitrarily small amount.”) 

We provide a precise definition on p 15 below. We’ll see that our definition is quite different from saying: “the constraint is satisfied with equality.” The definition we provide is consistent with what, by our estimate, roughly 99% of economists mean, when they say “a constraint is binding.”

Note the following important distinction between “satisfied with equality” and “binding.” In order to determine whether or not a constraint is satisfied with equality at a solution to an NPP, you \textit{have to specify a particular solution}. On the other hand, a constraint is either binding or not, regardless of which solution you select. For example consider the maximization problem,

$$\max_{x \in \mathbb{R}} u(x) = \max(x^3, 0) \quad \text{s.t.} \quad p \cdot x \leq 0,$$

Any scalar $$x \leq 0$$ solves this problem. However, the constraint $$x \leq 0$$ is satisfied with equality at only one of the solutions, i.e., at $$x = 0$$. On the other hand, the constraint $$x \leq 0$$ is binding, regardless of which solution you choose. Summarizing, being “satisfied with equality” is a property of a point, together a constraint; it has nothing to do with the objective function. Being “binding,” on the other hand, is a property of a constraint and an NPP; it has everything to do with the objective function.

---

2 Mathematical economics textbooks writers have to be more precise. And so, most of them, including Simon & Blume, de la Fuente, Varian, Berck and Sydaester and many others, use the word “binding” in a totally different way. They just define binding as: given an NPP and a point $$x^*$$ at which the NPP is maximized, a constraint is binding at $$x^*$$ if the constraint is satisfied with equality at this point. Note well, this is what they say, it is \textit{not} an acceptable definition for this class.
Returning to the Mantra... NB: you never, never, never worry about the second sentence of the mantra until after you’ve found a point that satisfies the first, i.e., until you’ve found a point $x^*$ that belongs to the nonnegative cone defined by the constraints that are satisfied with equality at $x^*$. Computer algorithms use the first sentence primarily: they go around the constraint set looking at points, checking which constraints are satisfied with equality, looking at the nonnegative cones that these constraints generate, and checking the first sentence. Only when they’ve found a point where the first sentence is satisfied do they bother with the second sentence.

Our definition of a “binding” constraint is restrictive in the sense that a badly behaved maximization problem may not have a maximum. Eg., maximize $f(x) = x$ s.t. $x < 1$. A problem of this kind does, however, have a supremum, defined in the footnote below. The technical definition is not important at this point: just think of the supremum of a set as being the “moral maximum” of the set, i.e., the only number that could possibly be the maximum if the set were to have something called a maximum. In these example above, the supremum of the set $\{f(x) : x < 1\}$ is clearly 1. In any non-linear programming problem, the “supremized” value of an objective function is a well-defined concept: it is the supremum of the values that the function can take on the constraint set. (This is a term you will not find in any math book.)

We say that a constraint is a binding constraint if it satisfies the following property: if you relax the constraint a little bit, then the “supremized” (“infimized”) value of the objective function will be increased (decreased). (This is a term you will find in even fewer math books.)

---

3 Definition: Let $S \subset \mathbb{R}$. A number $b \in \mathbb{R}$ is an upper bound for $S$ if $s \leq b$, for all $s \in S$. A number $b$ is the least upper bound for $S$ if $b$ is an upper bound for $S$ and if $b \leq b'$, for any upper bound $b'$ of $S$. The supremum of a set $S$ is now defined as the least upper bound for $S$, if $S$ has an upper bound, otherwise it is infinity.

4 If you are doing a constrained minimum problem, then this should be changed to infimized, where infimum is to minimum as supremum is to maximum.
We’ll now define both binding and linding precisely. Given an objective function \( f \) and a constraint set \( A \), define the \textit{maximized} (supremized) value of the problem \((f, A)\) to be the maximum (supremum) of the set \( \{f(x) : x \in A\} \). (Note that this maximum (supremum) lives in the \textit{range} not the domain.) Clearly, every problem \((f, A)\) for which \( f \) is bounded has a supremized value while lots of such problems won’t have a maximized value. Now consider the problem,

\[
\max f(x) \text{ s.t. } g^j(x) \oplus^j b^j, j = 1, \ldots m
\]

where for all \( j \), \( \oplus^j \in \{<, \leq\} \). (By replacing the usual \( \leq \) with \( \oplus^j \), we allow for the possibility that some of the constraints must be satisfied with strict equality and others with weak inequality.) Let \( M(b) \) (\( S(b) \)) denote the maximized (supremized) value of the problem \((f, A)\), where \( A = \{x : g^j(x) \oplus^j b^j, j = 1, \ldots m\} \). Define \( e^j \) by \( e_k^j = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \). (This is just the vector with 1 in the \( j \)'th component and zeroes elsewhere.) For a problem \((f, A)\) such that \( f \) is bounded, we now say that constraint \( j \) is \textit{binding} (linding) if there exists \( \bar{\epsilon} > 0 \) such that \( \forall \epsilon \in (0, \bar{\epsilon}), M(b + \epsilon e^j) > M(b) \). \( (S(b + \epsilon e^j) > S(b)) \).

Here are a few examples illustrating the relationship between lindingness and bindingness.

1. \( \max f(x) = x \text{ s.t. } x < 0 \). In this case, the single constraint, \( x < 0 \), is linding but not binding. (Notice that issue of whether or not a constraint is linding has \textit{nothing} to do with whether or not a solution (maximum) to the problem exists.)

2. \( \max f(x) = x \text{ s.t. } x \leq 0 \). In this case, the single constraint, \( x \leq 0 \), is again linding, since it’s binding, and bindingness is sufficient for lindingness.

3. \( \max f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \text{ s.t. } x < 0 \). In this case, the single constraint, \( x < 0 \), is both binding and linding, even though the constraint set is not closed.
Coming back to the mantra, there are two exceptions to the nonnegative cone rule. One arises only because words aren’t as precise as mathematical symbols. The other is substantive.

- (purely semantic) if the constrained max problem happens to have an unconstrained solution the mantra as stated above will typically not work. Here the problem is a matter of terminology: when we write the math version of the mantra, then it certainly works. the words of the mantra don’t exactly capture the mathematics.
- (substantive) the other exception is the so-called Constraint Qualification. More on it later.

There are several different versions of constrained optimization (Fig. 23): Each of the three below has a linear and a nonlinear version. Assume throughout the lecture that $f$ is quasi-concave and $g$ is quasi-convex.
(1) Maximize a function subject to an equality constraint. (i.e., solution has to lie on a line.)

\[
\max f(x) \text{ subject to } g(x) = b
\]

In other words, maximize \( f \) subject to the condition that \( x \) lies on a level set of the function \( g \). We’ll call \( f \) the objective function and \( g \) the constraint function.

(2) Maximize a function subject to an equality constraint and nonnegativity constraints

\[
\max f(x) \text{ subject to } g(x) = b; x \geq 0.
\]

In other words, maximize \( f \) subject to the condition that \( x \) lies on a level set of the function \( g \) and \( x \) is a nonnegative vector.

(3) Maximize a function subject to one or more inequality constraints.

- a single constraint: \( \max f(x) \text{ subject to } g(x) \leq b. \)
- multiple constraints: \( \max f(x) \text{ subject to } g(x) \leq b. \)

In other words, maximize \( f \) subject to the condition that \( x \) lies in a lower contour set or the intersection of several.

The difference between linear and nonlinear is whether the constraints are linear functions or not: i.e., if the level sets are straight lines, or not.

**Question:** Notice that we assume “≤”. What about greater than or equal to...?

**Answer:** If the original constraint is \( g(x) \geq b \), write it as \( -g(x) \leq -b \).

**Fact:** The first five cases (found below) are all special cases of the sixth. This is important: you really want to be able to fit everything into the same general framework.
• A linear function is just a special case of a nonlinear function.

• Write $g(x) = b$ as two constraints: $-g(x) \leq -b$; $g(x) \leq b$.

• What about nonnegativity constraints: the set $\{x \in \mathbb{R}^2 : x_i \geq 0\}$ is a lower contour set of a certain function, i.e., the function $g_i(x) = -x_i$; Plot $g_i(x)$ in 3D. So we can draw the conventional budget set as the intersection of three lower contour sets.

Aside: It’s useful here to observe the relationship between the budget set, which people are used to, and the lower contour set of the function $p \cdot x$. People find it hard to see where the gradient vector attached to the budget set comes from. That’s because they don’t think about the budget set as the lower contour set of a function.