Contents

1. Linear Algebra

1.1. Preliminary: Level Sets, upper and lower contour sets and Gradient vectors

Before we proceed with linear algebra, we need some concepts that we'll use to illustrate a key linear algebra concept.

You all presumably know what contour lines are on a map: lines joining up the points on the map that are at the same height above sea level. Same thing in math, except slightly different terms. Three terms that you need to know:

(1) **Level set:** A level set of a function f consists of all of the points *in the domain of* f at which the function takes a certain value. In other words, take any two points that belong to the same level set of a function f: this means that f assigns the same value to both points.
(2) **Upper contour set:** An upper contour set of a function f consists of all of the points *in the domain of* f at which the value of the function is *at least* a certain value. We talk about "the upper contour set of a function f corresponding to α", refering to the set of points to which f assigns the value at least α.
(3) **Lower contour set:** A lower contour set of a function f consists of all of the points *in the domain of* f at which the value of the function is *no more than* a certain value.

For example, consider Fig. 1 below. The level sets are indicated on the diagram by dotted lines. Very important fact that everybody gets wrong: the level sets are the lines on the horizontal plane at the bottom of the graph, NOT the lines that are actually on the graph. That is, the level sets are *points in the domain of the function* above which the function is the same height.

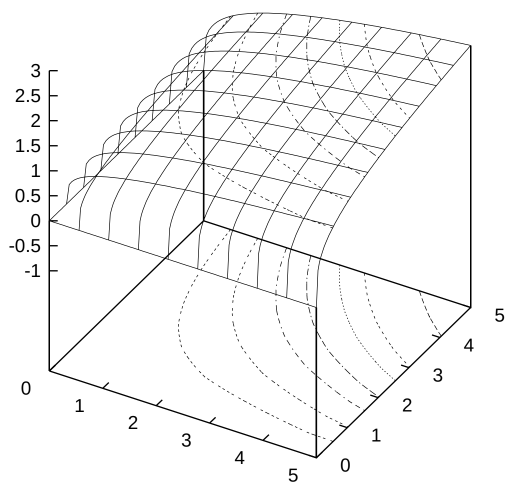

FIGURE 1. Level and contour sets of a concave function

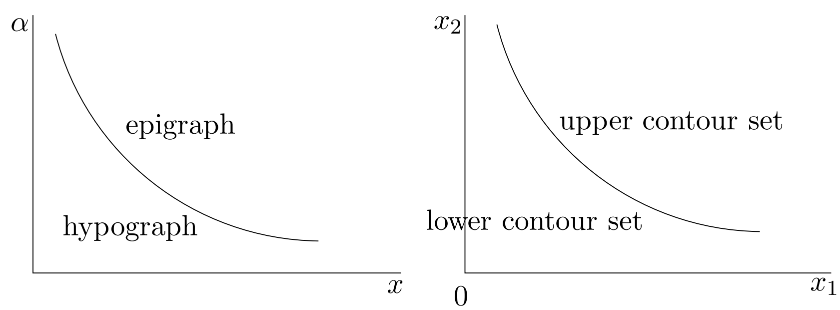

FIGURE 2. The graphs of one function and the level sets of another

Pick the first level set in the picture: suppose that the height of the function for every point on the level set is α. Notice that for every point above and to the right of this level set, the value of the function at this point is larger than α. Hence the *set* of points above and to the right of this level set is the upper contour set of the function corresponding to the value α.

The following is a source of endless confusion for everybody: compare the two curves in Fig. 2. The two curves are identical except for the labels. The interpretation of the curves is entirely different.

(1) On the left, we have the *graph* of a function of one variable; area NE of the line is the area above the graph, called the *epigraph* of the function; area SW of the line is the area below the graph, called the *hypograph* of the function;
(2) On the right, we have the *level set* of a function of *two* variables; area NE of the line is an upper contour set of the function; area SW of the line is an lower contour set of the function. In this case, the two-dimensional picture represents the domain of the function; the height of the function isn't drawn.

Where are the upper contour sets located in the left panel of the figure? Ans: pick α on the vertical axis. Find $x(\alpha)$ on the horizontal axis that's mapped to that point α. The interval $[0, x(\alpha)]$ is the upper contour set corresponding to α.

Some familiar economic examples of level sets and contour sets.

(1) level sets that you know by other names: indifference curves; isoprofit lines; budget line ($\boldsymbol{p} \cdot \mathbf{x} = y$). the production possibility frontier (this is the zero level set of the function $q - f(\mathbf{x})$).
(2) lower contour sets that you know by other names: budget *sets*; production possibility set;
(3) upper contour sets that you know by other names: think of the "region of comparative advantage" in an Edgeworth box: this is the intersection of the upper contour sets of the two traders' utility functions.

Some practice examples for level sets.

- What can we say about the level sets of a function $f : \mathbb{R} \to \mathbb{R}$ with no flat spots? Ans: A discrete (i.e., isolated) set of points.
- What are the level sets of a *concave* $f : \mathbb{R} \to \mathbb{R}$ variable function with no flat spots? How many points can be in a level set? Ans: At most two.
- Now consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ that has a strict local maximum at $\mathbf{x}^*$ (i.e., f is strictly higher at $\mathbf{x}^*$ than on a nbd). What can you say about the level set of the function through $\mathbf{x}^*$? Ans: The point $\mathbf{x}^*$ must be an *isolated* member of the level set. It's not necessarily the unique leement of the level set, but it's necessarily isolated.

<u>Vectors</u>: Recall that a vector in $\mathbb{R}^n$ is an ordered 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.

<u>Gradient vectors</u>: 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 $\mathbf{x}$, written as $\mathbf{f}'(\mathbf{x})$ or $\bigtriangledown f(\mathbf{x})$. Its components are the partial derivatives of the function f, evaluated at $\mathbf{x}$, i.e., $(f_1(\mathbf{x}), \cdots, f_n(\mathbf{x}))$

Example: $f(\mathbf{x}) = 2x_1x_2$, evaluated at $(2, 1)$, i.e., $\mathbf{f}'(2, 1) = (2x_2, 2x_1) = (2, 4)$. Draw the level set through $(2, 1)$, draw the gradient through the origin, lift it up and place its base at $(2, 1)$. The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a point in $\mathbb{R}^n$, and 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.

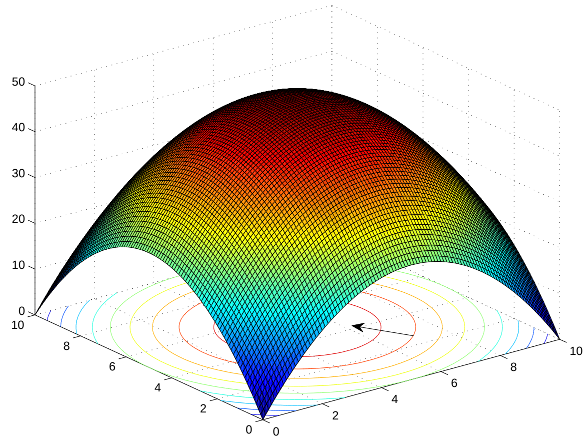

FIGURE 3. Level set and gradient vector through a point

The gradient vector points in the direction of steepest ascent: Consider Fig. 3. Let $\mathbf{x}$ denote the point in the domain where the base of the arrow touches the circle. The graph represents a nice symmetric mountain which you are currently about to scale. You are currently at the point $\mathbf{x}$. 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 $\mathbf{x}$ is the steepest way up? Answer is: the direction perpendicular to the slope of the level set. Draw an arrow pointing in this direction. Now the *gradient vector* of f at $\mathbf{x}$ 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 $\mathbf{x}$
- it points into the upper contour set. Note that this is *not* the same as saying that the head (or tip) of the vector lies inside of the upper contour set:; The arrow could pass thru the upper contour set, cross the level set again, and then pass into the lower contour set!
- 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.

1.2. **Vectors as arrows.**

Write vectors as arrows but the "real vector" is the location of the tip of the arrow. As noted above, it's important that in visual applications, we often draw vectors that don't have their base at the origin. E.g., the gradient vector at $\mathbf{x}$ is always drawn with its base at the point $\mathbf{x}$. Strictly speaking, you have to shift the arrow back to the origin in order to interpret it as a vector.

1.3. Vector operations

Note well that *all* of the following definitions assume implicitly that the vectors concerned all live in the same space, i.e., you can't add, subtract or multiply vectors unless they have the same number of elements.

Row and column vectors: doesn't make any difference whether the vector is written as a row or a column vector. Purely a matter of notational convenience.

For the purposes of this section, the *norm* of a vector is its Euclidean length: measure the arrow with a ruler.[1] Written $||\mathbf{x}|| = \sqrt{\sum_{k=1}^{n} x_k^2}$. Note that $||\mathbf{x}|| = d_2(\mathbf{x}, \mathbf{0})$.

The *inner product* of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is the sum of the products of the components. That is, $\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k$. When I think of the inner product of $\mathbf{x}$ and $\mathbf{y}$, I think of $\mathbf{x}$ as a row vector and of $\mathbf{y}$ as a column vector; purely a convention.

It is hard to visualize what $\mathbf{x} \cdot \mathbf{y}$ looks like. Look at a picture of $\mathbf{x}$ and $\mathbf{y}$ and say whether $\mathbf{x} \cdot \mathbf{y}$ is positive, negative, zero. Answer is given by the angle between the two vectors.

- acute angle means $\mathbf{x} \cdot \mathbf{y}$ is positive.
- obtuse angle means $\mathbf{x} \cdot \mathbf{y}$ is negative.
- ninety degree angle means $\mathbf{x} \cdot \mathbf{y}$ is zero.

Theorem: $\mathbf{a} \cdot \mathbf{u} = ||\mathbf{a}|| \, ||\mathbf{u}|| cos(\theta)$, where θ is the angle between $\mathbf{a}$ and $\mathbf{u}$. Note the beauty of cos: doesn't matter whether you look at the big angle between the vectors or the little one, get the same answer!

In Fig. 4, rank the inner products $\mathbf{x} \cdot \mathbf{a}$, $\mathbf{x} \cdot \mathbf{b}$ and $\mathbf{x} \cdot \mathbf{c}$. Answer: all the vectors are the same length, so that the only thing that determines the inner product is the angle between them. Hence $\mathbf{x} \cdot \mathbf{a} > \mathbf{x} \cdot \mathbf{b} > \mathbf{x} \cdot \mathbf{c}$.

Application: a fact that we'll study in the Calculus section is that for small vectors $\mathbf{dx}$, $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x}) + \bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx}$. Just take this on faith for the moment.

- Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Draw $\mathbf{x}$ and $\mathbf{dx}$ in $\mathbb{R}^2$, then add them to get $\mathbf{x} + \mathbf{dx}$. Now think about $f(\mathbf{x} + \mathbf{dx})$: is it bigger or smaller than $f(\mathbf{x})$?
- First answer is graphical. (See Fig. 5.) Draw $\mathbf{x}$ and a level set through $\mathbf{x}$. Now draw 3 small vectors $\mathbf{dx}$ starting from $\mathbf{x}$.

[1] I'd never have said this in the analysis section, because there are *lots* of different possible norms. But in this section, a certain lack of obsessiveness is forgivable.

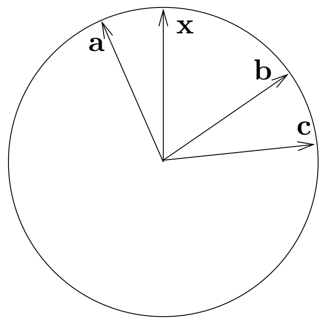

FIGURE 4. Inner Products.

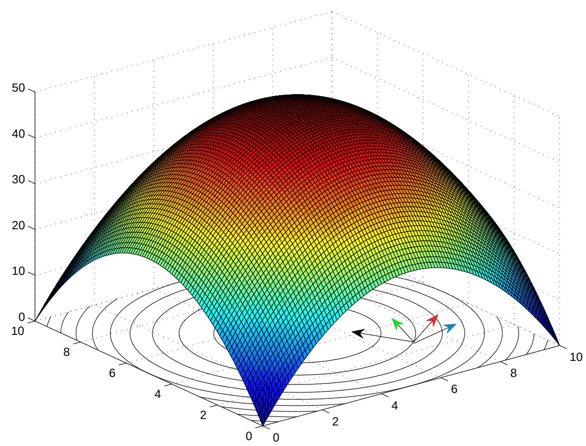

FIGURE 5. Level set and gradient vector through a point

- The green arrow, whose head is located at $\mathbf{x} + \mathbf{dx}$, belongs to the upper contour set of f corresponding to $f(\mathbf{x})$. That is, by definition of the upper contour set of f corresponding to $f(\mathbf{x})$, $f(\mathbf{x}+\mathbf{dx}) > f(\mathbf{x})$. Which $\mathbf{dx}$'s point into this particular upper contour set? A *necessary* condition for $\mathbf{x} + \mathbf{dx}$ to belong to this upper contour set is that $\mathbf{dx}$ makes an acute angle with $\bigtriangledown(\mathbf{x})$. But this acute angle condition is not *sufficient* for $\mathbf{x} + \mathbf{dx}$ to belong to this upper contour set. Why not?
- The blue arrow, whose head is located at $\mathbf{x}+\mathbf{dx}$, belongs to the lower contour set of f corresponding to $f(\mathbf{x})$, That is, by definition of the lower contour set of f corresponding to $f(\mathbf{x})$, $f(\mathbf{x} + \mathbf{dx}) < f(\mathbf{x})$. A *sufficient* condition for $\mathbf{x} + \mathbf{dx}$ to belong to this lower contour set is that $\mathbf{dx}$ makes an obtuse angle with $\bigtriangledown(\mathbf{x})$. But this obtuse angle condition is not *necessary* for $\mathbf{x} + \mathbf{dx}$ to belong to this upper contour set. Why not?
- *What's the source of the asymmetry here?*.
  - i) the acute angle condition is *necessary* but not *sufficient* for $f(\mathbf{x} + \mathbf{dx})$ to belong to the *upper* contour set corresponding to...
  - ii) the obtuse angle condition is *sufficient* but not *necessary* for $f(\mathbf{x} + \mathbf{dx})$ to belong to the *lower* contour set corresponding to...

Answer: the gradient vector doesn't tell the whole story, it only gives us approximate information: It turns out (we'll learn this formally later) that for some $\lambda \in [0,1]$,

$$f(\mathbf{x}+\mathbf{dx}) \quad \text{is exactly equal to} \quad f(\mathbf{x}) \quad + \quad \bigtriangledown f(\mathbf{x})\cdot\mathbf{dx} \quad + \quad 0.5\mathbf{dx}'H(\mathbf{x}+\lambda\mathbf{dx})\mathbf{dx}.$$

We'll learn that since f is a strictly concave function, the second term is always negative—technically H is a *negative definite matrix*—and hence there's a horserace going on in case i) but not in case ii).

- if $\mathbf{dx}$ is tangent to the level set (the red arrow)—i.e., $\mathbf{dx}$ and $\bigtriangledown f(\mathbf{x})$ are *orthogonal* to each other, i.e., the angle between them is a 90 degree or right angle–then $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x})$ i.e., f is *approximately* flat in this direction, Note, however, that actually the red arrow points into the *lower* contour set. The function *decreases* when you move along the level set. Why? The approximation $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x})+\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}$ provides us with useful information *only* when $\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx} \neq 0$. When $\bigtriangledown f\cdot\mathbf{x} = 0$, you have to consider a more accurate approximation $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x})+\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}+0.5\mathbf{dx}'H(\mathbf{x})\mathbf{dx}$. We'll see later that while for the red $\mathbf{dx}$, the $\bigtriangledown f$ term is zero, the last term is negative, hence this $\mathbf{dx}$ points into the lower contour set

- Now observe that, provided $\mathbf{dx}$ is sufficiently small, you get the same answer as the graphical one when you use the fact that $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x})+\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}$ and apply the cos formula to $\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}$. Answer depends on the angle between $\bigtriangledown f(\mathbf{x})$ and $\mathbf{dx}$.
  - if angle is acute, then $f(\mathbf{x}+\mathbf{dx}) > f(\mathbf{x})$.
  - if angle is 90°, then $f(\mathbf{x}+\mathbf{dx}) \approx f(\mathbf{x})$.
  - if angle is obtuse, then $f(\mathbf{x}+\mathbf{dx}) < f(\mathbf{x})$.

  This verifies that the gradient of a function at $\mathbf{x}$ points into the upper contour set of the function at $\mathbf{x}$, and that the gradient is perpendicular to the level set.

1.4. Linear Combinations, Linear Independence, Linear Dependence and Cones.

Defn: $\mathbf{x} \in \mathbb{R}^n$ is a *linear combination* of a set of m vectors $\{\mathbf{v}^1, \ldots\mathbf{v}^k, \ldots\mathbf{v}^m\}$ in $\mathbb{R}^n$ if there exists a vector $\mathbf{t} \in \mathbb{R}^m$ such that $\mathbf{x} = \sum_{k=1}^m t_k\mathbf{v}^k$; in words, if $\mathbf{x}$ can be written as the sum of scalar multiples of the original vectors $\mathbf{v}^k$'s.

Example: If $\mathbf{v}^1, \mathbf{v}^2 \in \mathbb{R}^2$ point in the same (or opposite) direction, then the linear combinations of these vectors all lie on the same line. If not, then any point in $\mathbb{R}^2$ can be written as a linear combination of $\mathbf{v}^1$ and $\mathbf{v}^2$.

Defn: $\mathbf{x} \in \mathbb{R}^n$ is a *nonnegative linear combination* of a set of m vectors $\{\mathbf{v}^1, \ldots\mathbf{v}^k \ldots, \mathbf{v}^m\}$ in $\mathbb{R}^n$ if there exists a vector $\mathbf{t} \in \mathbb{R}^m_+$ such that $\mathbf{x} = \sum_{k=1}^n t_k\mathbf{v}^k$. I.e., the coefficients all have to be nonnegative.

Defn: The *nonnegative (positive) cone* defined by a set of vectors $\{\mathbf{v}^1, \ldots\mathbf{v}^k \ldots, \mathbf{v}^m\}$ is the set of all nonnegative (positive) linear combinations of these vectors.

Note that

(1) if you have two vectors in $\mathbb{R}^2$ that aren't collinear, the difference between the nonnegative cone and the positive cone is that the "edges" of the cone aren't included in the positive cone but are included in the nonnegative cone.

(2) if you have two vectors in $\mathbb{R}^2$ that are collinear, and the second is a *negative* multiple of the first, i.e., you have the vectors $\mathbf{x}$ and $\alpha\mathbf{x}$, with $\alpha < 0$, then the positive and the nonnegative cones are identical and consist of the entire line through these vectors

Convex combinations: what's the difference between one of these and a linear combination? Ans: the set of convex combinations of *two* vectors is the line segment between them, which is a subset of the nonnegative cone defined by these two vectors. The set of convex combinations of *three* vectors is a triangle embedded in a plane.

Defn: $\mathbf{x} \in \mathbb{R}^n$ is a *convex combination* of a set of vectors $\{\mathbf{v}^1, ...\mathbf{v}^k..., \mathbf{v}^m\}$ if there exists a vector $\mathbf{t} \in \mathbb{R}^m_+$ such that $\sum_{k=1}^m t_k = 1$ and $\mathbf{x} = \sum_{k=1}^m t_k\mathbf{v}^k$. I.e., the coefficients have to be nonnegative and sum to one.

The next concept we're going to define is *linear independence*, which is a fundamental concept in linear algebra. I'm going to give you two definitions, one that's intuitive, but not quite correct. The other is not intuitive at all, but has the redeeming virtue of being correct.

Friendly but not quite correct defn: a set of vectors $\{\mathbf{v}^1, ...\mathbf{v}^k..., \mathbf{v}^m\}$, $\mathbf{v}^k \in \mathbb{R}^n$, is a *linear independent set* if no one of them can be written as a linear combination of all the others.

Unfriendly but correct defn: a set of vectors $\{\mathbf{v}^1, ...\mathbf{v}^k..., \mathbf{v}^m\}$ is a *linear independent set* if for all $\mathbf{t} \in \mathbb{R}^m$, $\sum_{k=1}^m t_k\mathbf{v}^k = 0$ implies $\mathbf{t} = 0$.

Defn: a set of vectors $\{\mathbf{v}^1, ..., \mathbf{v}^m\}$ is a *linear dependent set* if it is not a linear independent set.

The *only* difference between my friendly defn and the unfriendly formal definition is that the set $\{0\}$ is linear independent by my definition and linear dependent by the formal definition. To see that it is linear independent according to my definition, note that if 0 is the only element of the set, then, trivially, you can't write 0 as a lin comb of other vectors in the set, because there aren't any other vectors in the set to write it as a linear combination of. To see that it is linear dependent by the formal definition, let $t = 1$, and note that $t \times 0 = 0$, but $t \neq 0$. So the test for linear independence fails. It's a useful exercise to check that the two definitions are equivalent for any set of vectors *other than* the singleton set zero.

Note that the empty set is a linear independent set, since it trivially satisfies the definition of linear independence.

Examples:

- Can you construct a linear independent set of vectors in which one of the vectors is zero?
- What's the largest set of linearly independent 2-vectors you can have?
- What's the largest set of linearly independent 3-vectors you can have?