PSC 403

– Vector Structure of $\mathbb{R}^n$, or Linear Algebra

– Recall the discussion on pp. 22-24 of the class notes about ordered pairs and n-tuples. An ordered pair (x,y) is a way of keeping track of two things and their order. Thus, as long as we know the horizontal axis corresponds to the first component and the vertical axis corresponds to the second, we can use (x,y) to denote a point on the Cartesian plane. Turning this around, we can think of the Cartesian plane as the set of all ordered pairs of real numbers.

– An ordered n-tuple, say $(x_1, x_2, \cdots x_n)$ keeps track of n things and their order. We define n-dimensional Euclidean space as the set of ordered n-tuples of real numbers. We denote it by $\mathbb{R}^n$. (Note that $\mathbb{R}^1 = \mathbb{R}$, and $\mathbb{R}^2$ is the Cartesian plane.) Elements of $\mathbb{R}^n$ are vectors.

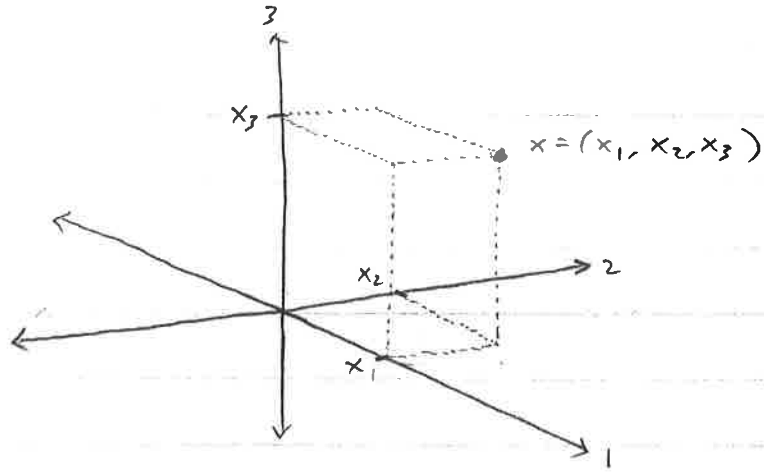
We may write $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, arranging $x_1 \cdots x_n$ vertically.

– If the number n is understood, we may write simply x for $(x_1, x_2 \cdots x_n)$, y for $(y_1, y_2 \cdots y_n)$, etc.

– Points in $\mathbb{R}^1$ can be depicted as points on the real line, and points in $\mathbb{R}^2$ can be depicted as points in the Cartesian plane. Points in $\mathbb{R}^3$ can also

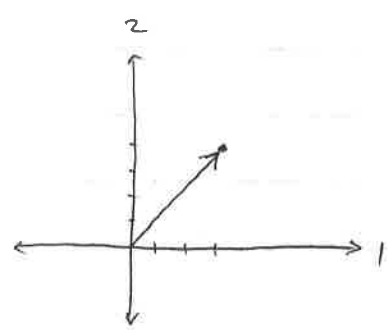
PSC 403

be depicted graphically, if we draw one axis pointing "out from the page."



Points in $\mathbb{R}^4$ or higher can't be depicted graphically. Fortunately, $\mathbb{R}^3$ will give us all the intuition we need for higher dimensional analysis.

- Though points in $\mathbb{R}^n$ are really points, it can be helpful sometimes to think of them as arrows, i.e., $x \in \mathbb{R}^n$ is an arrow pointing from the origin $(0, 0 \ldots 0)$ to $(x_1, x_2 \ldots x_n)$. In $\mathbb{R}^2$, if $x = (3, 4)$, we can think of it as follows:



- Note: We will write 0 instead of $(0, 0 \ldots 0)$.

- Given x and y in $\mathbb{R}^n$, we define <u>vector addition</u> as follows:

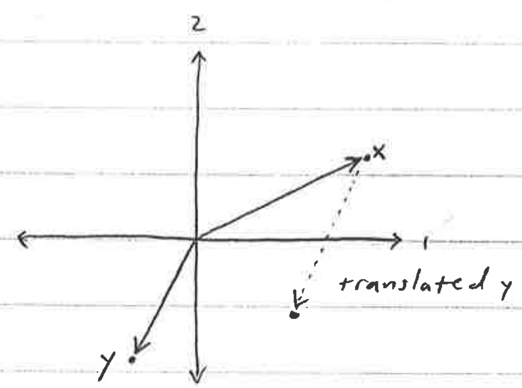
$$x+y = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n).$$

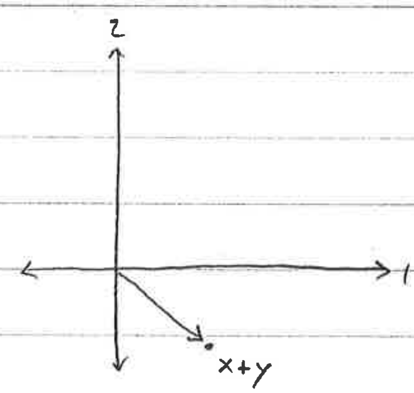
- Given $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$, define <u>scalar multiplication</u> as

$$\alpha x = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)$$

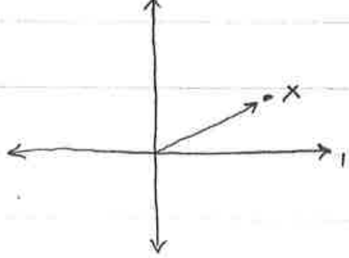
(Note: the term "scalar" refers to a real number.)

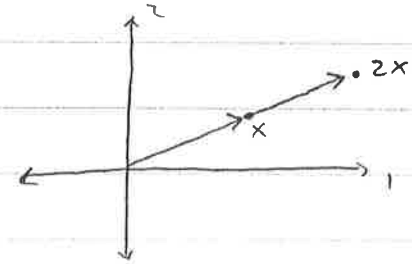
- Geometrically, $x+y$ is the result of "translating" y so that the tail of the arrow starts at the point x; the head of (translated) y gives us the head of $x+y$.





- And αx is the result of "scaling" x by a factor of α. If $\alpha = 2$, we have the situation below:





- If $\alpha < 0$ then the direction of x is reversed.

- The following properties of vector addition and scalar multiplication in $\mathbb{R}^n$ are easily checked.

(i) $x + y = y + x$

(ii) $(x+y)+z = x + (y+z)$

(iii) $x + 0 = x$

(iv) $x + (-1)x = 0$

(v) $\alpha(\beta x) = (\alpha\beta)x$

(vi) $1 \cdot x = x$

(vii) $\alpha(x+y) = (\alpha x) + (\alpha y)$

(viii) $(\alpha + \beta)x = (\alpha x) + (\beta y)$

(Note: We write $-x$ instead of x.)

optional comment

- As a matter of interest, it turns out that $\mathbb{R}^n$ is, in a sense, the only set with these properties. More precisely, if X has an addition operation, a scalar multiplication operation, and a zero vector satisfying (i) through (viii) and it is "finite-dimensional" then, for some n, X is "equivalent" to $\mathbb{R}^n$.

- Given $\alpha_1, \alpha_2 \cdots \alpha_m \in \mathbb{R}$ and $x_1, x_2 \cdots x_m \in \mathbb{R}^n$, we call $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m$ a <u>linear combination</u> of $x_1, x_2 \cdots x_m$.

- For any k between 1 and n, we define e^k as

$$e^k = (0 \ldots 0, \underset{\uparrow\, k^{th} \text{ coordinate}}{1}, 0 \ldots 0)$$

If $n = 3$ then $e^1 = (1,0,0)$, $e^2 = (0,1,0)$, and $e^3 = (0,0,1)$. There are called the <u>unit coordinate vectors</u>.

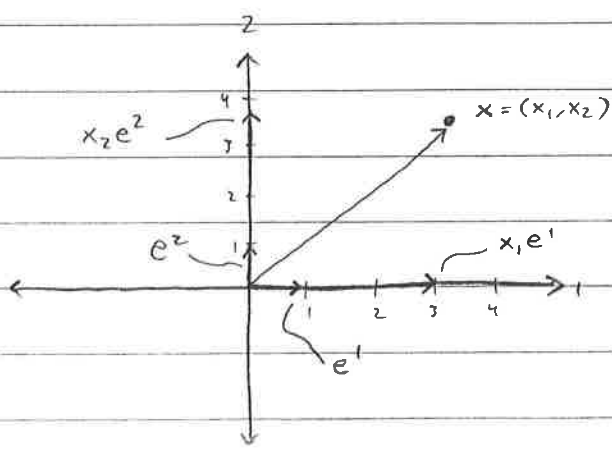
- Note that every $x \in \mathbb{R}^n$ can be written as a linear combination of $e^1, e^2, \ldots e^n$. How? Letting $x = (x_1, x_2 \ldots x_n)$, it is clear that

$$x = x_1 e^1 + x_2 e^2 + \cdots + x_n e^n.$$

Moreover, this linear combination is <u>unique</u>. By this, I mean that

$$x = \alpha_1 e^1 + \alpha_2 e^2 + \cdots + \alpha_n e^n \Rightarrow \alpha_1 = x_1, \ \ldots, \ \alpha_n = x_n.$$

See below:



PSC 403

- Given a set $\{x^1, x^2, \dots x^m\} \subseteq \mathbb{R}^n$ of vectors (note that the indexes are now superscripts) the <u>span</u> of $\{x^1, \dots x^m\}$ is the set, denoted $\text{span}\{x^1 \dots x^m\}$, of linear combinations of $x^1, x^2 \dots x^m$, i.e.,

$$\text{span}\{x^1 \dots x^m\} = \{ \alpha_1 x^1 + \dots + \alpha_m x^m \mid \alpha_1, \alpha_2 \dots \alpha_m \in \mathbb{R} \}.$$

- Note: (i) $\text{span}\{e^1, e^2, \dots e^n\} = \mathbb{R}^n$
  (ii) $\text{span}\{x\}$ is the line through x.
  (iii) $\text{span}\{x^1, \dots x^m\} \subseteq \text{span}\{x^1, \dots x^m, x^{m+1}\}$.

- Letting $n=2$, $\text{span}\{(2,1), (3,4)\} = \mathbb{R}^2$. To see this, take any $x = (x_1, x_2) \in \mathbb{R}^2$ and solve

$$\alpha_1 (2,1) + \alpha_2 (3,4) = (x_1, x_2),$$

or,

$$(\alpha_1 2, \alpha_1) + (\alpha_2 3, \alpha_2 4) = (x_1, x_2),$$

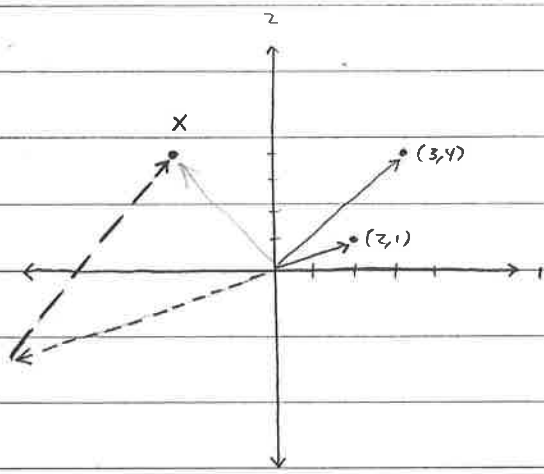
or,

$$(\alpha_1 2 + \alpha_2 3, \ \alpha_1 + \alpha_2 4) = (x_1, x_2),$$

or,

$$2\alpha_1 + 3\alpha_2 = x_1$$
$$\alpha_1 + 4\alpha_2 = x_2$$

for α_1 and α_2. We get $\alpha_1 = \frac{4x_1 - 3x_2}{5}$ and $\alpha_2 = \frac{2x_2 - x_1}{5}$. You can check that $x = \alpha_1(2,1) + \alpha_2(3,4)$. Note that this linear combination is <u>unique</u>.

- 

To see how this works geometrically, take x as to the left. We multiply $(2,1)$ by some $\alpha_1 < -1$ to get the $\leftarrow$--- vector, and we multiply $(3,4)$ by some $\alpha_2 > 1$? to get the $\nearrow$ vector. Adding these, we get x.

- So far we have seen $\text{span}\{e^1 \ldots e^n\} = \mathbb{R}^n$ and $\text{span}\{(2,1),(3,4)\} = \mathbb{R}^2$, and more: not only does $\{e^1,\ldots,e^n\}$ span $\mathbb{R}^n$, but every vector in $\mathbb{R}^n$ is a $\underset{\sim\sim}{\text{unique}}$ linear combination of $e^1,\ldots,e^n$, and likewise for $\{(2,1),(3,4)\}$ and $\mathbb{R}^2$.

- This is not always the case: note that $\text{span}\{(1,0),(0,1),(2,0)\} = \mathbb{R}^2$, but now $x = (x_1, x_2)$ can be written as a linear combination in many ways:

$$x = x_1(1,0) + x_2(0,1)$$

and

$$x = \frac{x_1}{2}(2,0) + x_2(0,1)$$

and

$$x = \frac{x_1}{3}(1,0) + \frac{x_1}{3}(2,0) + x_2(0,1),$$

etc.

- An important concept in understanding this is the following. A set $\{x^1, x^2, \ldots x^m\} \subseteq \mathbb{R}^n$ is <u>linearly dependent</u> if one element can be written as a linear combination of the others, i.e., there exist $\alpha_1, \alpha_2, \ldots, \alpha_m$, not all zero, such that

$$\alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_m x^m = 0.$$

The set is <u>linearly independent</u> if it is not linearly dependent.

- Note that $\{e^1, \ldots, e^n\}$ is linearly independent, and so is $\{(2,1), (3,4)\}$. To check the latter claim, suppose

$$\alpha_1 (2,1) + \alpha_2 (3,4) = 0,$$

or,

$$2\alpha_1 + 3\alpha_2 = 0$$
$$\alpha_1 + 4\alpha_2 = 0.$$

Then $\alpha_1 = -4\alpha_2$, so $2\alpha_1 + 3\alpha_2 = -8\alpha_2 + 3\alpha_2 = -5\alpha_2 = 0$ implies $\alpha_2 = 0$, and then $\alpha_1 = 0$.

- Geometrically, this is obvious. A set of two vectors, say x and y, is linearly dependent iff there is some α such that $x = \alpha y$, i.e., they lie on the same line.

Theorem ② If $\{x^1, x^2, \cdots, x^m\}$ is a linearly independent set of m vectors and $y^1, y^2, \cdots, y^k \in \text{Span}\{x^1, x^2, \cdots, x^m\}$ is a set of $k > m$ vectors, then $\{y^1, y^2, \cdots, y^k\}$ is linearly dependent. (971021T 10:40)

PSC 403

- But $\{(1,0), (0,1), (2,0)\}$ is linearly dependent, since $(1,0) = \frac{1}{2}(2,0)$. This turns out to be the key distinction.

- Theorem: ① The set $\{x^1, \ldots, x^m\} \subseteq \mathbb{R}^n$ is linearly independent iff every $x \in \text{span}\{x^1 \ldots x^m\}$ is a unique linear combination of $x^1 \ldots x^m$, i.e.,

$$\left.\begin{array}{l} x = \alpha_1 x^1 + \cdots + \alpha_m x^m \\ \text{and} \quad x = \beta_1 x^1 + \cdots + \beta_m x^m \end{array}\right\} \Rightarrow \begin{array}{l} \alpha_1 = \beta_1, \ \alpha_2 = \beta_2, \\ \ldots, \ \alpha_m = \beta_m. \end{array}$$

- Proof: Suppose $\{x^1 \ldots x^m\}$ is linearly independent, $x = \alpha_1 x^1 + \cdots + \alpha_m x^m = \beta_1 x^1 + \cdots + \beta_m x^m$, but $\alpha_k \neq \beta_k$ for some k. Then

$$(\alpha_k - \beta_k) x^k = (\beta_1 - \alpha_1) x^1 + \cdots + (\beta_{k-1} - \alpha_{k-1}) x^{k+1} + (\beta_{k+1} - \alpha_{k+1}) x^{k+1} + \cdots + (\beta_m - \alpha_m) x^m.$$

Since $\alpha_k - \beta_k \neq 0$,

$$x^k = \sum_{j \neq k} \frac{(\beta_j - \alpha_j)}{(\alpha_k - \beta_k)} x_j,$$

contradicting linear independence.

Now suppose $\{x^1 \ldots x^m\}$ is linearly dependent, so there exist $\alpha_1, \ldots \alpha_m$, not all zero, s.t.

$$\alpha_1 x^1 + \dots + \alpha_m x^m = 0.$$

Say $\alpha_k \neq 0$. But then x^k can be written as

$$x^k = 1 \cdot x^k + \sum_{j \neq k} 0 \cdot x^j$$

and

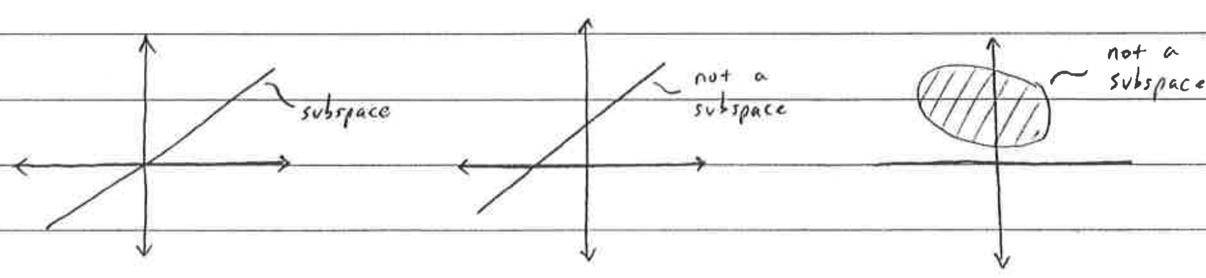
$$x^k = 0 \cdot x^k + \sum_{j \neq k} \frac{\alpha_j}{\alpha_k} x^j,$$

which are clearly distinct linear combinations. //

- A set $X \subseteq \mathbb{R}^n$ is a <u>subspace</u> if, for all $m \in \mathbb{N}$, all $\alpha_1, \dots, \alpha_m \in \mathbb{R}$, and all $x^1, \dots, x^m \in X$, $\alpha_1 x^1 + \dots + \alpha_m x^m \in X$. In words, if every linear combination of elements in X is in X.

- Check: If X and Y are subspaces, so is $X \cap Y$.

- In $\mathbb{R}^2$, $\{0\}$ and $\mathbb{R}^2$ itself are subspaces, and, for every $x \in \mathbb{R}^2$, the line through 0 and x is a subspace — and there are no others.



PICTURE

- In $\mathbb{R}^3$, $\{0\}$, $\mathbb{R}^3$, lines, and <u>planes</u> through 0 are subspaces.

- If $X = \text{span}\{x^1, \dots, x^m\}$ and $\{x^1, \dots, x^m\}$ is linearly independent then we call $\{x^1, \dots, x^m\}$ a <u>basis</u> for X.

- Thus, given a basis for X, we can write elements of X as unique linear combinations of the basis vectors.

- As we have seen, $\{e^1, \dots, e^n\}$ is a basis for $\mathbb{R}^n$, and $\{(2,1), (3,4)\}$ is a basis for $\mathbb{R}^2$. What is a basis for the line through 0 and x? $\{x\}$. Note that $\{(2,1), (3,4)\}$ and $\{e^1, e^2\}$ are both bases for $\mathbb{R}^2$ (so a set can have more than one basis), and they have the same number of elements. This is no coincidence.

- Theorem:
  - (i) X is a linear subspace iff it has a basis.
  - (ii) If $\{x^1, \dots, x^m\}$ has m distinct elements and is a basis for X and $\{y^1, \dots, y^k\}$ has k distinct elements and is a basis for X, then $m = k$. ?
  - (iii) If $\{x^1, \dots, x^m\}$ is a basis for X and $\text{span}\{y^1, \dots y^k\} = X$ then $k \geq m$ (assuming $x^1 \dots x^m$ are distinct and $y^1 \dots y^k$ are distinct).

- Proof omitted.

- Thus, if X is a subspace, it has a basis; and all its bases have the same number, say d, elements. We call d the <u>dimension</u> of X.

- Since $\mathbb{R}^n$ has a basis with n elements, namely $\{e^1, \dots, e^n\}$, it has dimension equal to n (as you would expect.

- As a consequence of part (iii) of the preceding theorem, it takes at least d vectors to span a d-dimensional subspace.

- Application to systems of m linear equations in n unknowns:

homogeneous

non-homogeneous

$m<n$, $m=n$, $m>n$

- Theorem: If $\{x^1 \ldots x^m\}$ is a set of m distinct vectors and $\{y^1 \ldots y^k\} \subseteq \text{span}\{x^1 \ldots x^m\}$ is a set of $k > m$ distinct vectors then $\{y^1 \ldots y^k\}$ is linearly dependent.

- Proof: Since $\{y^1 \ldots y^k\} \subseteq \text{span}\{x^1 \ldots x^m\}$, we can write

$$y^1 = \alpha_1^1 x^1 + \cdots + \alpha_m^1 x^m$$
$$\vdots$$
$$y^m = \alpha_1^m x^1 + \cdots + \alpha_m^m x^m$$
$$y^{m+1} = \alpha_1^{m+1} x^1 + \cdots + \alpha_m^{m+1} x^m$$
$$\vdots$$
$$y^k = \alpha_1^k x^1 + \cdots + \alpha_m^k x^m.$$

Without loss of generality, we can assume $\alpha_1^1 \neq 0$. Then

$$x^1 = +\frac{1}{\alpha_1^1} y^1 - \frac{\alpha_2^1}{\alpha_1^1} x^2 - \cdots - \alpha_m^1/\alpha_1^1 \, x^m,$$

and $\text{span}\{x^1 \ldots x^m\} \subseteq \text{span}\{y^1, x^2 \ldots x^m\}$. (Check this.) Thus, $y^2 \in \text{span}\{y^1, x^2 \ldots x^m\}$, i.e.,

$$y^2 = \beta_1 y^1 + \beta_2 x^2 + \cdots + \beta_m x^m.$$

If $\beta_2 = \beta_3 = \cdots = \beta_m = 0$ then y^2 is a linear combination of y^1, and we are done. Otherwise, suppose without loss of generality that $\beta_2 \neq 0$. Then

$$x^2 = -\beta_1/\beta_2\, y^1 + 1/\beta_2\, y^2 - \beta_3/\beta_2\, x^3 - \dots - \beta_m/\beta_2\, x^m,$$

and $\operatorname{span}\{y^1, x^2, \dots x^m\} \subseteq \operatorname{span}\{y^1, y^2, x^3 \dots x^m\}$. Thus, $y^3 \in \operatorname{span}\{y^1, y^2, x^3 \dots x^m\}$, i.e.,

$$y^3 = \gamma_1 y^1 + \gamma_2 y^2 + \gamma_3 x^3 + \dots + \gamma_m x^m.$$

If $\gamma_3 = \gamma_4 = \dots = \gamma_m = 0$ then y^3 is a linear combination of y^1 and y^2, and we are done. Otherwise, suppose without loss of generality that $\gamma_3 \neq 0$. The argument proceeds as above (formally, we should use an induction argument here), yielding

$$\operatorname{span}\{x^1 \dots x^m\} \subseteq \operatorname{span}\{y^1\, x^2 \dots x^m\}$$

$$\subseteq \operatorname{span}\{y^1, y^2, x^3 \dots x^m\} \subseteq \dots \subseteq \operatorname{span}\{y^1, y^2 \dots y^m\}$$

after m steps. But then $y^{m+1} \in \operatorname{span}\{y^1 \dots y^m\}$, so $\{y^1 \dots y^{m+1}\}$, and therefore $\{y^1 \dots y^k\}$, is linearly dependent.

//

PSC 403

$$\alpha_1 x^1 + \dots + \alpha_m x^m = 0.$$

Say $\alpha_k \neq 0$. But then x^k can be written as

$$x^k = 1 \cdot x^k + \sum_{j \neq k} 0 \cdot x^j$$

and

$$x^k = 0 \cdot x^k + \sum_{j \neq k} \frac{\alpha_j}{\alpha_k} x^j,$$

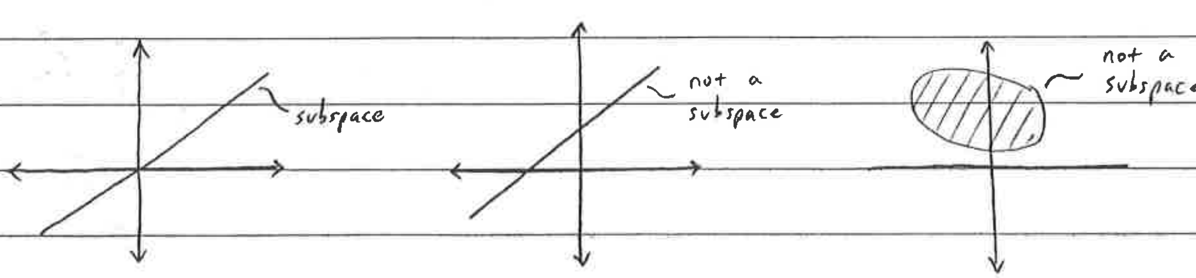
which are clearly distinct linear combinations. //

- A set $X \subseteq \mathbb{R}^n$ is a (linear) subspace if, for all $m \in \mathbb{N}$, all $\alpha_1, \dots, \alpha_m \in \mathbb{R}$, and all $x^1, \dots, x^m \in X$, $\alpha_1 x^1 + \dots + \alpha_m x^m \in X$. In words, if every linear combination of elements in X is in X.

Note: $\text{span}\{x^1, \dots, x^m\}$ is a subspace.

- Check: If X and Y are subspaces then $X \cap Y$ is a subspace.

- In $\mathbb{R}^2$, $\{0\}$ and $\mathbb{R}^2$ itself are subspaces, and, for every $x \in \mathbb{R}^2$, the line through 0 and x is a subspace — and there are no others.



PICTURE

- In $\mathbb{R}^3$, $\{0\}$, $\mathbb{R}^3$, lines, and planes through 0 are subspaces.

- Given a set X, if $X = \text{span}\{x^1, \ldots, x^m\}$ and $\{x^1, \ldots, x^m\}$ is linearly independent then we call $\{x^1, \ldots, x^m\}$ a <u>basis</u> for X.

- Thus, given a basis for X, we can write elements of X as unique linear combinations of the basis vectors.

- As we have seen, $\{e^1, \ldots, e^n\}$ is a basis for $\mathbb{R}^n$, and $\{(2,1), (3,4)\}$ is a basis for $\mathbb{R}^2$. What is a basis for the line through 0 and x? $\{x\}$. Note that $\{(2,1), (3,4)\}$ and $\{e^1, e^2\}$ are both bases for $\mathbb{R}^2$ (so a set can have more than one basis), and they have the same number of elements. This is no coincidence.

- Theorem:
  (i) For all subspaces $X \subseteq \mathbb{R}^n$ and all $x^1, \ldots x^m \in X$, (subspace) $\{x^1 \ldots x^m\}$ is a basis for X iff it is a maximal linearly independent subset of X, i.e., it is linearly independent and there is no linearly independent subset of X with more elements.
  (ii) Any two bases of subspace X have the same number of elements.
  (iii) X is a subspace iff it has a basis.

$X \subseteq \mathbb{R}^n$

Take any $x^1 \in X$ such that $x^1 \neq 0$

Then $\{x^1\}$ is linearly independent.

If $span\{x^1\} = X$ we're done

If not, there is some $x^2 \in X$ such that $x^2 \notin span\{x^1\}$, then $\{x^1, x^2\}$ is linearly independent.

Why? Suppose $\exists \alpha_1, \alpha_2$, such that $\alpha_1 x^1 + \alpha_2 x^2 = 0$

If $\alpha_2 = 0$, $x^1 = 0$, so $\alpha_2 \neq 0$

Then $x^2 = -\frac{\alpha_1}{\alpha_2} x_1$, i.e. $x^2 \in Span\{x^1\}$

If $span\{x^1, x^2\} = X$ we're done

Else, take $x^3 \notin span\{x^1, x^2\}$, $\{x^1, x^2, x^3\}$ is linearly independent

Suppose $\exists \alpha_1, \alpha_2, \alpha_3$ such that $\alpha_1 x^1 + \alpha_2 x^2 + \alpha_3 x^3 = 0$

This can't continue more than n steps.

If it did, say n+1, we would have $\{x^1, x^2, \cdots x^n, x^{n+1}\}$ linearly independent.

But $\{e^1, \cdots, e^n\}$ is a maximal linearly independent set.

971028T9:40

Proposition

i) The rank of $\{x^1, x^2, \cdots, x^m\} \leq m$, i.e., the dimension of $span\{x^1, x^2, \cdots, x^m\}$ is $\leq m$

Proof: Let $\{y^1, y^2, \cdots, y^k\}$ be a maximal linearly independent subset of $\{x^1, x^2, \cdots, x^m\}$

If $x^k \neq 0$, then $\{x^k\}$ is linearly independent.

Since $\{x^1, x^2, \cdots, x^m\}$ is finite, there has to be a biggest linearly independent subset.

Claim. $Span\{y^1, y^2, \cdots, y^k\} \supseteq Span\{x^1, x^2, \cdots, x^m\}$

① Each $x^k \in Span\{y^1, y^2, \cdots, y^k\}$

Suppose not. Then $\{y^1, y^2, \cdots, y^k, x^k\}$ is a bigger linearly independent subset of $\{x^1, x^2, \cdots, x^m\}$

To see that $\{y^1, y^2, \cdots, y^k, x^k\}$ is linearly independent, suppose $\exists \alpha_1, \alpha_2, \cdots, \alpha_k, \alpha_k \in \mathbb{R}$ not all zero such that

$(\alpha_1 y^1 + \alpha_2 y^2 + \cdots \alpha_k y^k) + \alpha_k x^k = 0$ $\alpha_k \neq 0$

but then $x^k \in Span\{y^1, y^2, \cdots, y^k\}$

② Since $\{x^1, x^2, \cdots, x^m\} \subseteq span\{y^1, y^2, \cdots, y^k\}$ and $Span\{y^1, y^2, \cdots, y^k\}$ is a subspace,

$Span\{x^1, x^2, \cdots, x^m\} \subseteq span\{y^1, y^2, \cdots, y^k\}$

- Thus, if X is a subspace, it has a basis; and all its bases have the same number, say d, elements. We call d the <u>dimension</u> of X.

- Since $\mathbb{R}^n$ has a basis with n elements, namely $\{e^1, \dots, e^n\}$, it has dimension equal to n (as you would expect).

PROPOSITION: The rank of $\{x^1 \dots x^m\}$ is $\leq m$. If X has dimension d and $\{x^1 \dots x^m\} \subseteq X$ has rank r then $r \leq d$. If $r = d$ then $\text{span}\{x^1 \dots x^m\} = X$. So if $\{x^1 \dots x^m\} \subseteq \mathbb{R}^n$ is a linearly independent set, $m \leq n$.

- As a consequence of part (ii) of the preceding theorem, it takes at least d vectors to span a d-dimensional subspace.

- The <u>rank</u> of $\{x^1, x^2 \dots x^m\}$ is the dimension of $\text{span}\{x^1, x^2 \dots x^m\}$.

- Application to systems of linear equations. Consider m equations in n unknowns, $x_1, \dots x_n$, as below,

$$
\begin{aligned}
a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n &= y_1 \\
a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n &= y_2 \\
\vdots \qquad\qquad\qquad\qquad & \quad \vdots \\
a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{mn} x_n &= y_m,
\end{aligned}
$$

where $a_{k,l}$ is a fixed real number, $k = 1 \dots m$, $l = 1 \dots n$.

PSC 403

- We can think of $(a_{1\ell}\ a_{2\ell}\ \cdots\ a_{m\ell})$ as a vector in $\mathbb{R}^m$. Writing this vertically for each $\ell = 1 \ldots n$, the above system of equations is

$$\begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} x_2 + \cdots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

- Thus, the system of equations has a solution iff $(y_1\, y_2 \ldots y_m)$ is a linear combination of $(a_{11}, a_{2,1} \ldots a_{m,1}), \ldots (a_{1n}, a_{2n}, \ldots a_{mn})$. And this means

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \text{span}\left\{ \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}.$$

- If $(y_1, y_2 \cdots y_m) = 0$ we call the system of equations <u>homogeneous</u>.

- Proposition: For a homogeneous system of linear equations,
  (i) a solution always exists, namely $x_1 = x_2 = \cdots x_n = 0$.
  (ii) the set of solutions is a subspace.
  (iii) consequently, either there is exactly one solution (the zero solution) or there are infinitely.

PSC 403

- Note: There is exactly one solution iff the vectors $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$ are linearly independent. (In fact, this is the definition of linear independence.)

- Note: Think about what it means for the system of equations to always have a solution, meaning, to have a solution for all $(y_1, y_2 \cdots y_m)$. It means that every point in $\mathbb{R}^m$ is in the span of the "column vectors" $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$. That is,

$$\mathbb{R}^m \subseteq \operatorname{span}\left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}.$$

- Note that every linear combination of the column vectors is in $\mathbb{R}^m$, so the above inclusion becomes

$$\mathbb{R}^m = \operatorname{span}\left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\},$$

or equivalently,

$$\operatorname{rank}\left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\} = m.$$

- Note: If the system always has exactly one solution, then $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$ are linearly independent and span $\mathbb{R}^m$. Therefore, there are m of them, so $m=n$. It can also be shown that the row vectors, $(a_{11} \cdots a_{1n}), \cdots, (a_{m1} \cdots a_{mm})$, are linearly independent!

PSC 403

- Proposition: For *any* system of linear equations,
  (i) rank $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\} < m$ iff there is some $(y_1, y_2 \cdots y_m)$ for which there is no solution.
  (ii) $n < m$ implies there is some $(y_1, y_2 \cdots y_m)$ for which there is no solution.
  (iii) rank $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\} = m$ if and only if, for all $(y_1, y_2 \cdots y_m)$, a solution exists.
  (iv) $n = m$ and $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}$ linearly independent if and only if, for all $(y_1, y_2 \cdots y_m)$, there is a *unique* solution to the system of equations.

- Note how (ii) follows from (i): rank $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\} \leq n$ and $n < m$ yield the antecedent of part (i).

- Proposition: Given a system of m linear equations with n unknowns and arbitrary $(y_1 \cdots y_m)$, let $X \subseteq \mathbb{R}^n$ denote the set of solutions to the corresponding system of homogeneous equations,

$$\begin{aligned} a_{11} x_1 + \cdots + a_{1n} x_n &= 0 \\ \vdots \qquad\qquad \vdots \quad &\ \ \vdots \\ a_{m1} x_1 + \cdots + a_{mn} x_n &= 0. \end{aligned}$$

Let $z \in \mathbb{R}^n$ denote *any* solution to the original system of equations.

(i) X is a subspace.

(ii) w is a solution to the original system of equations iff there exists $x \in X$ such that $w = z + x$.

- Thus, the subspace of solutions to the system of homogeneous equations and a single solution of the original system are enough to determine all solutions of the original system.

- Example:

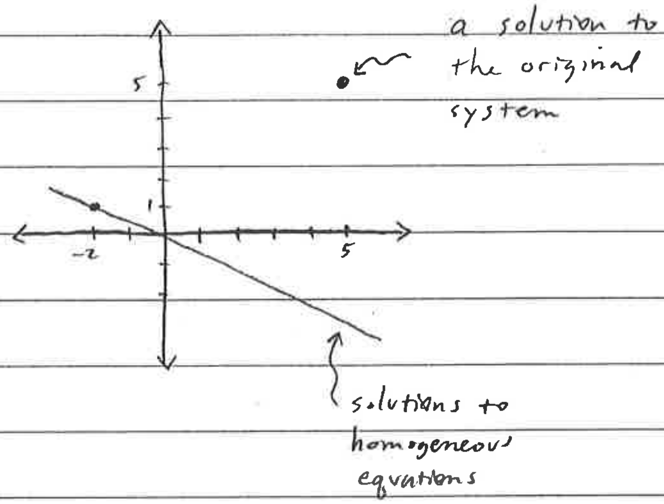
$$2x_1 + 4x_2 = 30$$
$$3x_1 + 6x_2 = 45.$$

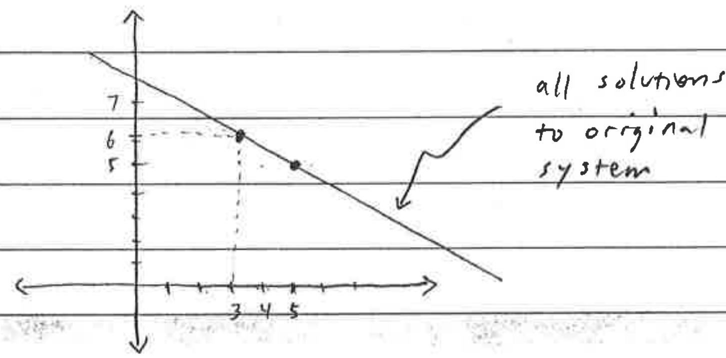
Note that $z = (5,5)$ is a solution to this system of equations. And (x_1, x_2) solves the corresponding system of homogeneous equations,

$$2x_1 + 4x_2 = 0$$
$$3x_1 + 6x_2 = 0,$$

iff $x_1 = -2x_2$ iff (x_1, x_2) is on the line connecting $(-2,1)$ to the origin. Graphically, the situation is as below.

PSC 403





Geometry of $\mathbb{R}^n$

- How can we go about finding the solutions to a system of equalities? That's where matrix representations become useful.

Matrix Algebra

- A matrix is a rectangular array of real numbers. If it has m rows and n columns, we might denote it

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

- Geometry of $\mathbb{R}^n$

  - The length, or norm, of a vector $x \in \mathbb{R}^n$ is denoted $\|x\|$ and defined by

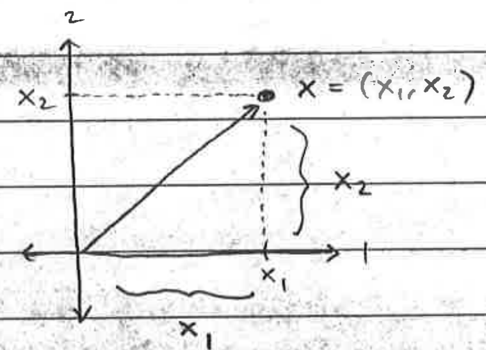
$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

(This means the positive square root of $x_1^2 + x_2^2 + \cdots + x_n^2$.)

- If $\|x\| = 1$, x is a unit vector.
- Note that $\frac{1}{\|x\|} x$ is always a unit vector.

  - Note that the idea of "norm" generalizes the idea of absolute value. That is, if $x \in \mathbb{R}^1$ then $\|x\| = |x|$.

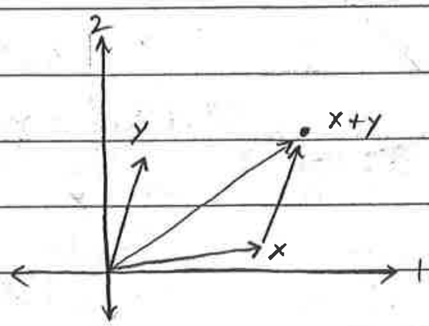
  - Also, norm in $\mathbb{R}^2$ is consistent with the "Pythagorean Theorem."



$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

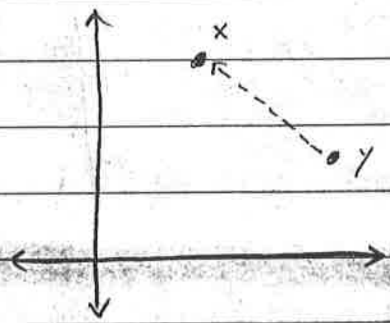
  - Proposition: For all $\alpha \in \mathbb{R}$ and all $x, y \in \mathbb{R}^n$,
    - (i) $\|\alpha x\| = |\alpha| \, \|x\|$.
    - (ii) $\|x\| = 0$ iff $x = 0$.
    - (iii) $\|x + y\| \leq \|x\| + \|y\|$.

  - Part (iii) is called the "triangle inequality." It formalizes the notion that "the shortest distance between two points is a straight line."

– 

Clearly, $\|x+y\| \le \|x\| + \|y\|$.

– The <u>distance</u> between two vectors x and y is $\|x-y\|$.
Note that $\|x-y\| = \|y-x\|$.



By definition,
$y + (x-y) = x$,
so $x-y$ is the dashed vector to the left.

– The <u>inner product</u> (or <u>dot product</u>) of $x, y \in \mathbb{R}^n$ is denoted $\langle x, y \rangle$ (or $x \cdot y$) and defined by

– Generalizes multiplication of real numbers.

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

– Proposition: For all $\alpha, \beta \in \mathbb{R}$ and all $x, y, z \in \mathbb{R}^n$,

(i) $\langle x, x \rangle = \|x\|^2$.

(ii) $\langle x, y \rangle = \langle y, x \rangle$.

(iii) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$.

- Note that part (iii) can be broken into two parts:

(a) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$

(b) $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.

- Let's check these parts. To see (a), note that

$$\langle \alpha x, y \rangle = (\alpha x_1) y_1 + (\alpha x_2) y_2 + \cdots + (\alpha x_n) y_n$$

$$= \alpha \left(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \right)$$

$$= \alpha \langle x, y \rangle.$$

To see (b), note that

$$\langle x+y, z \rangle = (x_1 + y_1) z_1 + (x_2 + y_2) z_2 + \cdots + (x_n + y_n) z_n$$

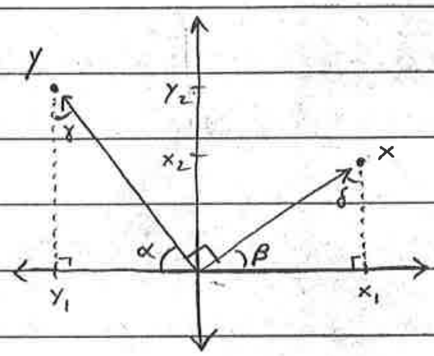
$$= x_1 z_1 + y_1 z_1 + x_2 z_2 + y_2 z_2 + \cdots + x_n z_n + y_n z_n$$

$$= \langle x, z \rangle + \langle y, z \rangle.$$

- Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if $\langle x, y \rangle = 0$.

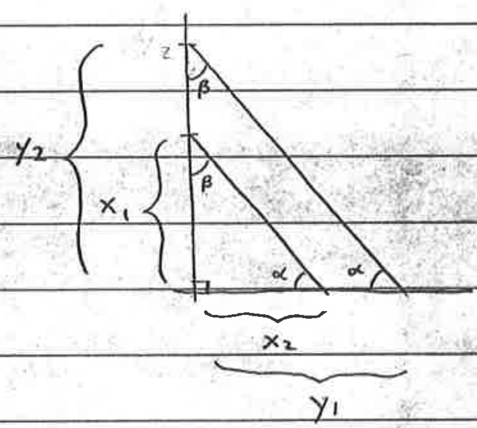
- This generalizes the usual notion of perpendicular-ness. (What's the word for this?) That is, for $x, y \in \mathbb{R}^2$, $\langle x, y \rangle = 0$ iff the vectors, viewed as arrows, are perpendicular.

- To see this, take any $x, y \in \mathbb{R}^2$. If the angle between x and y is $90°$,



$\alpha + \gamma = \alpha + \beta = 90$ implies $\gamma = \beta$ and $\beta + \delta = \alpha + \beta = 90$ implies $\alpha = \delta$. Therefore, the triangles formed to the left are similar.

- Translating the triangles, there is some $c > 0$ such that



$$y_2 = c x_1$$
$$y_1 = -c x_2.$$

Then

$$x_1 y_1 + x_2 y_2 = -c x_1 x_2 + c x_1 x_2 = 0.$$

Therefore, $\langle x, y \rangle = 0$.

– Proposition: For all $x, y \in \mathbb{R}^n$, if α is the (acute) angle between x and y then

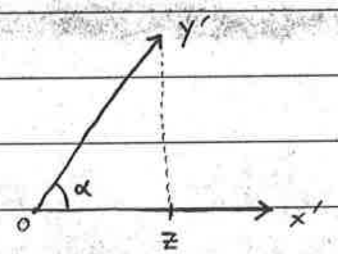
$$\langle x, y \rangle = \|x\| \, \|y\| \cos\alpha .$$

– I'll sketch the argument for $\mathbb{R}^2$. Take any $x, y \in \mathbb{R}^2$ and define

$$x' = \frac{1}{\|x\|} x \quad \text{and} \quad y' = \frac{1}{\|y\|} y .$$

Note that x' and y' are unit vectors.

–



Rotate x' and y', preserving the angle between them, so that x' is horizontal. Drop a line down to x'. Call the point where it hits the x' arrow z. Note that $z = \|z\| x'$, since x' is a unit vector. And, since x' and y' are unit vectors, $\cos\alpha = \|z\|$. Now, from $z = y + (z - y)$ and $\langle z - y', x' \rangle = 0$,

$$① \ \langle z, x' \rangle = \langle y' + (z - y'), x' \rangle = \langle y', x' \rangle + \langle z - y', x' \rangle = \langle x', y' \rangle .$$

Since $z = \|z\| x'$, $② \ \langle z, x' \rangle = \langle \|z\| x', x' \rangle = \|z\| \langle x', x' \rangle .$

Then $\|x'\|=1$ implies ③ $\langle x',x'\rangle=1$. Combining ①, ②, and ③, $\|z\|=\langle x',y'\rangle$. Therefore, $\cos\alpha=\langle x',y'\rangle$. Finally, note that

$$\cos\alpha=\langle x',y'\rangle=\left\langle \frac{1}{\|x\|}x,\frac{1}{\|y\|}y\right\rangle=\frac{1}{\|x\|\,\|y\|}\langle x,y\rangle,$$

which tells us $\langle x,y\rangle=\|x\|\,\|y\|\cos\alpha$.

- In particular, if x and y are unit vectors then $\langle x,y\rangle=\cos\alpha$. As we have seen, $\langle x,y\rangle=0$ exactly when $\cos\alpha=0$, i.e., x and y are orthogonal. Note that $\langle x,y\rangle$ is highest when $\cos\alpha=1$, i.e., $\alpha=0$. In other words, when x and y are unit vectors, $\langle x,y\rangle$ is maximized when $x=y$.

- Proposition (The Cauchy-Schwartz Inequality): $\langle x,y\rangle\le\|x\|\,\|y\|$.

- Proposition (the Triangle Inequality): $\|x+y\|\le\|x\|+\|y\|$.

- Proof: Note that, using the Cauchy-Schwartz Inequality,

$$\begin{aligned}\|x+y\|^2&=\langle x+y,x+y\rangle=\langle x,x\rangle+2\langle x,y\rangle+\langle y,y\rangle\\&\le\langle x,x\rangle+2\|x\|\,\|y\|+\langle y,y\rangle=\|x\|^2+2\|x\|\,\|y\|+\|y\|^2\\&=(\|x\|+\|y\|)^2.\end{aligned}$$

Taking square roots, $\|x+y\|\le\|x\|+\|y\|$. //

PSC 403

- A vector $x \in \mathbb{R}^n$ can be thought of as a matrix with n rows and one column $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ or as a matrix with one row and m columns $(x_1 \cdots x_n)$, depending on whether we write it vertically or horizontally.

- Given two matrices, A and B, of the same dimensions, say $m \times n$ (the first number is the number of rows, the second is the number of columns), define

$$A+B = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix}.$$

- Given a $m \times n$ matrix A and an $n \times 1$ matrix $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, i.e., a vector in $\mathbb{R}^n$, define

$$Ax = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} (a_{11} \cdots a_{1n}) \cdot (x_1 \cdots x_n) \\ (a_{21} \cdots a_{2n}) \cdot (x_1 \cdots x_n) \\ \vdots \\ (a_{m1} \cdots a_{mn}) \cdot (x_1 \cdots x_n) \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

Thus, Ax is a $m \times 1$ matrix, i.e., a vector in $\mathbb{R}^m$.

PSC 403

- Note that the system of linear equations,

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ \vdots \qquad\qquad \vdots \quad &\quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m \end{aligned}$$

can then be compactly written $Ax=y$, where

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \text{ and } y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

- Given an $m \times n$ matrix A and an $n \times k$ matrix B, define <u>matrix multiplication</u> by

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & & a_{11}b_{1k} + \cdots + a_{1n}b_{nk} \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1k} + \cdots + a_{2n}b_{nk} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & & a_{m1}b_{1k} + \cdots + a_{mn}b_{nk} \end{pmatrix}.$$

Thus, AB is a $m \times k$ matrix. This definition is more easily digested if we break B up into "column vectors," $b^1, b^2 \ldots b^k$, where

$$b^1 = \begin{pmatrix} b_{11} \\ \vdots \\ b_{n1} \end{pmatrix}, \quad b^2 = \begin{pmatrix} b_{12} \\ \vdots \\ b_{n2} \end{pmatrix}, \quad \ldots \quad b^k = \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix}.$$

We can write $B = \begin{pmatrix} b^1 & b^2 & \ldots & b^k \end{pmatrix}$. Then

$$AB = \begin{pmatrix} Ab^1 & Ab^2 & \cdots & Ab^k \end{pmatrix},$$

premultiply postmultiply

where Ab^1 is just "multiplication" of A by a column vector, defined on p. 132. Recall that, for all $j = 1 \dots k$, Ab^j is a $m \times 1$ matrix (vector). Since AB has k such columns, its dimension is $m \times k$.

- A $m \times n$ matrix is square if $m = n$.

- Matrix multiplication, defined above, shares some similarities with multiplication of numbers. Firstly, for all n, there is an $n \times n$ identity matrix I, defined by

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ 0 & 0 & \cdots & & 1 \end{pmatrix},$$

such that

① For all $m \times n$ matrices A,

$$AI = A$$

and

② For all $n \times m$ matrices B,

$$IB = B.$$

Secondly, it satisfies the following laws

(i) $(A+B)+C = A+(B+C)$

(ii) $(AB)C = A(BC)$

(iii) $A+B = B+A$

(iv) $A(B+C) = AB + AC$

(v) $(A+B)C = AC + BC$.

- For example,

$$\begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \left(\begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$$

$$= \begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}.$$

And

$$\left(\begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 3 & 8 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} 6 & -3 \\ -2 & 2 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\left(\begin{pmatrix} 3 & 8 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -2 & 2 \end{pmatrix}\right)$$

$$\begin{pmatrix} 6+0 & 16+5 \\ -3+0 & -8+6 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -2 & 2 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 18-16 & -9+16 \\ -2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 21 \\ -3 & -2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -2 & 2 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 2 & 5 \\ -1 & 6 \end{pmatrix}\begin{pmatrix} 2 & 7 \\ -2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 36-42 & -18+42 \\ -18+4 & 9-4 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 4-10 & 14+10 \\ -2-12 & -7+12 \end{pmatrix}$$

$$\begin{pmatrix} -6 & 24 \\ -14 & 5 \end{pmatrix} = \begin{pmatrix} -6 & 24 \\ -14 & 5 \end{pmatrix},$$

confirming (i) and (ii).

- Matrix multiplication is different from multiplication of numbers in several important ways. First, it is not generally commutative, i.e.,

$$AB = BA$$

is not generally true. Indeed, it would not even make sense to write this unless A and B were square matrices, say $n \times n$. But even then it is not generally true: for example,

$$\begin{pmatrix} 6 & 8 \\ -1 & -2 \end{pmatrix}\begin{pmatrix} 3 & -5 \\ 1 & 2 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 3 & -5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 6 & 8 \\ -1 & -2 \end{pmatrix}$$

$$\begin{pmatrix} 18+8 & -30+16 \\ -3-2 & 5-4 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 18+5 & 24+10 \\ 6-2 & 8-4 \end{pmatrix}$$

$$\begin{pmatrix} 26 & -14 \\ -5 & 1 \end{pmatrix} \neq \begin{pmatrix} 23 & 34 \\ 4 & 4 \end{pmatrix}.$$

Second, an arbitrary matrix A may not have a "multiplicative inverse," i.e., there may be no matrix B such that $AB = I$ or $BA = I$.

- If $AB = I$, we call B a right inverse of A. If

$CA = I$, we call C a left inverse of A.

- Proposition (Thm 8.5 and Lemma 8.1 of Simon & Blume):
  If B is a right inverse of A and C is a left inverse of A, $B = C$.

- Proof: Note that $CA = I$. "Post-multiplying" both sides by B, $(CA)B = IB = B$. By associativity, $C(AB) = B$. Since $AB = I$, $C = CI = C(AB) = B$. //

- Thus, if A has a left and right inverse, they are one and the same, denoted A^{-1}. We call this the inverse of A, and we say A is invertible.

- If A is invertible, the system $Ax = y$ of linear equations always has a unique solution, namely, $A^{-1}y$. To see this, note that

$$A(A^{-1}y) = (AA^{-1})y = Iy = y,$$

so it is indeed a solution. Also, suppose x is some solution. Then $Ax = y$. "Pre-multiplying" by A^{-1}, we get

$$x = Ix = (A^{-1}A)x = A^{-1}(Ax) = A^{-1}y.$$

- Proposition: A is invertible iff, for all y, the system $Ax=y$ has a unique solution.

- Proof: We just proved the $\Rightarrow$ direction. Now suppose $Ax=y$ always has a solution. In particular, $Ax=e^i$, where $e^i=\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}$, has a solution, say $c^i=\begin{pmatrix}c^i_1\\ \vdots\\ c^i_n\end{pmatrix}$.

That is, $Ac^i=e^i$ for $i=1\ldots m$. Letting $C=(c^1\ldots c^m)$ denote the $n\times m$ matrix with columns c^i, we have

$$AC=A(c^1\ldots c^m)=(Ac^1\cdots Ac^m)=(e^1\ldots e^m)=I.$$

Therefore, C is a right inverse. The existence of a left inverse is a little trickier. I noted on p.128 that $Ax=y$ always having a unique solution implied the row vectors of A are linearly independent. Then the "transpose" of A, denoted A^T (see Simon & Blume, p.157), also has a right inverse, B, i.e., $A^TB=I$. But then $B^TA=I$, so B^T is a left inverse of A. Thus, A is invertible. //

- Corollary: Let A be $m\times n$. Then A is invertible iff the column vectors of A are a basis for $\mathbb{R}^m$ (which implies $m=n$, i.e., A is square).

PSC 403

- If A is invertible, we can actually calculate A^{-1}. Once that's done, the solution is $A^{-1}y$. Solving for A^{-1} can be quite difficult. See Simon & Blume, pp. 167-169, for a description of a very elegant technique, using "Gauss-Jordan elimination," for finding the inverse of A.

(omitting elementary row operations, Gauss-Jordan elimination)

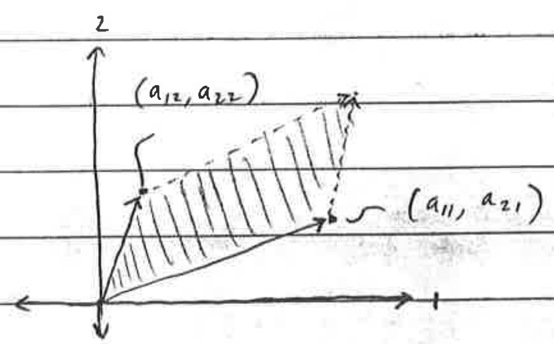
- We will focus on a different technique for solving $Ax=y$ that is common in economics and political science, called "Cramer's rule." To define this, however, I need to talk briefly about determinants.

digression on determinants ↓

- Determinants and Cramer's Rule, Simon + Blume Ch 9

- Given a 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, the determinant of A is defined as $a_{11}a_{22} - a_{21}a_{12}$.

- Geometrically, this is the area of the trapezoid formed below. Note that this area is zero exactly when the column vectors of A lie on the same line — when they are linearly dependent. Thus, we have a simple test to see when two vectors in $\mathbb{R}^2$ are



PSC 403

linearly dependent. (Of course, this is easy to tell anyway...) Moreover, it tells us how "close" to being linearly dependent they are: if the determinant of A is close to zero then the vectors are close to dependent.

- Remarkably, we can extend the definition of determinant to any $n \times n$ matrices, still giving us a test of linear independence.

- Given A, an $n \times n$ matrix, the <u>determinant</u> of A, denoted $\det A$, or $|A|$, is defined as follows. When $n=2$, $\det A = a_{11}a_{22} - a_{12}a_{21}$. When $n>2$, we need some more notation. Let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and j^{th} column. Then

$$\det A = \sum_{j=1}^{n} a_{1j} (-1)^{1+j} \det A_{1j}$$

$$= \underbrace{a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} - \cdots}_{n \text{ terms}}$$

The definition is "recursive" in the following sense: it tells us $\det A$ for $n=2$; then we can use this to figure out $\det A$ for $n=3$; and then $n=4$, and so on.

PSC 403

- Example:

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}\det\begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix} - a_{12}\det\begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}$$

$$+ a_{13}\det\begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix}$$

$$= a_{11}(a_{12}a_{23} - a_{22}a_{13}) - a_{12}(a_{11}a_{33} - a_{31}a_{13})$$
$$+ a_{13}(a_{11}a_{32} - a_{31}a_{12})$$

- In the definition, we fixed $i=1$ and let j vary between 1 and n. In fact, we can fix i at anything; or we can fix j at anything and let i vary from 1 to n. In general, we can fix i and define

$$\det A = \sum_{j=1}^{n} a_{ij}\,(-1)^{i+j}\det A_{ij},$$

or we can fix j and define

$$\det A = \sum_{i=1}^{n} a_{ij}\,(-1)^{i+j}\det A_{ij}.$$

Which row or column you use is solely a matter of convenience.

- Theorem: $\det A = 0$ iff $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}$ is linearly dependent.

- Example: I will expand this determinant along the second column.

$$\det\begin{pmatrix} 3 & 1 & 4 & 0 \\ 2 & 0 & 2 & 4 \\ 1 & 0 & 3 & 2 \\ 3 & 1 & 0 & 1 \end{pmatrix} = -\begin{vmatrix} 2 & 2 & 4 \\ 1 & 3 & 2 \\ 3 & 0 & 1 \end{vmatrix} + 0 - 0 + \begin{vmatrix} 3 & 4 & 0 \\ 2 & 2 & 4 \\ 1 & 3 & 2 \end{vmatrix}$$

$$= -\left[3\begin{vmatrix} 2 & 4 \\ 3 & 2 \end{vmatrix} + \begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix}\right] + \left[-4\begin{vmatrix} 3 & 4 \\ 1 & 3 \end{vmatrix} + 2\begin{vmatrix} 3 & 4 \\ 2 & 2 \end{vmatrix}\right]$$

$$= -\left[3(4-12) + (6-2)\right] + \left[-4(9-4) + 2(6-8)\right]$$

$$= 24 - 4 - 20 - 4$$

$$= -4.$$

Therefore (assuming my algebra is right), the vectors $\{(3,2,1,3), (1,0,0,1), (4,2,3,0), (0,4,2,1)\}$ are linearly independent.

- Let's summarize some equivalences we've obtained. The following are equivalent statements for an $m \times n$ matrix A.

① The system $Ax = y$ has a unique solution for all $y \in \mathbb{R}^m$.

② $n = m$ and $\operatorname{rank}\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\} = n$.

③ The collection $\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}$ is a basis for $\mathbb{R}^m$.

④ A is invertible.

⑤ $n=m$ and $|A| \neq 0$.

- When any of these conditions holds, we call A <u>non-singular</u>.

- We can use the determinant to construct a mechanical method for inverting a non-singular matrix A. (See Simon + Blume, p.194.)

- More importantly, perhaps, there is a method, called <u>Cramer's Rule</u>, for finding the solution of $Ax=y$ for any non-singular A. Let B_j denote the $n \times n$ matrix obtained from A by replacing the jth column by $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Then the unique solution to $Ax=y$ is $x=(x_1, x_2 \cdots x_n)$, where Then $X=(X_1 \cdots X_n)$ is given by $X_j = \frac{|B_j|}{|A|}$

$$x_j = \frac{\det B_j}{\det A},$$

i.e., $$x = \left(\frac{\det B_1}{\det A}, \frac{\det B_2}{\det A}, \ldots, \frac{\det B_n}{\det A} \right).$$

- We'll see with an example that this method isn't exactly painless. But it's nice to have an algorithm for computing solutions.

PSC 403

– Example: Solve

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ 12x_1 + 2x_2 - 3x_3 &= 5 \\ 3x_1 + 4x_2 + x_3 &= -4. \end{aligned}$$

This is $Ax = y$, where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 0 \\ 5 \\ -4 \end{pmatrix}.$$

So

$$x_1 = \frac{\begin{vmatrix} 0 & 1 & 1 \\ 5 & 2 & -3 \\ -4 & 4 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 12 & 5 & -3 \\ 3 & -4 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{vmatrix}}, \quad x_3 = \frac{\begin{vmatrix} 1 & 1 & 0 \\ 12 & 2 & 5 \\ 3 & 4 & -4 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{vmatrix}}.$$

We have three determinants to compute.

$$\begin{vmatrix} 1 & 1 & 1 \\ 12 & 2 & -3 \\ 3 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 2 & -3 \\ 4 & 1 \end{vmatrix} - \begin{vmatrix} 12 & -3 \\ 3 & 1 \end{vmatrix} + \begin{vmatrix} 12 & 2 \\ 3 & 4 \end{vmatrix}$$

$$= (2 + 12) - (12 + 9) + (48 - 6)$$

$$= 14 - 21 + 42$$

$$= 35.$$

PSC 403

$$\begin{vmatrix} 0 & 1 & 1 \\ 5 & 2 & -3 \\ -4 & 4 & 1 \end{vmatrix} = -\begin{vmatrix} 5 & -3 \\ -4 & 1 \end{vmatrix} + \begin{vmatrix} 5 & 2 \\ -4 & 4 \end{vmatrix} = -(5-12) + (20+8) = 35.$$

$$\begin{vmatrix} 1 & 0 & 1 \\ 12 & 5 & -3 \\ 3 & -4 & 1 \end{vmatrix} = \begin{vmatrix} 5 & -3 \\ -4 & 1 \end{vmatrix} + \begin{vmatrix} 12 & 5 \\ 3 & -4 \end{vmatrix} = (5-12) + (-48-15) = -70$$

$$\begin{vmatrix} 1 & 1 & 0 \\ 12 & 2 & 5 \\ 3 & 4 & -4 \end{vmatrix} = \begin{vmatrix} 2 & 5 \\ 4 & -4 \end{vmatrix} - \begin{vmatrix} 12 & 5 \\ 3 & -4 \end{vmatrix} = \overset{-28}{(-8-20)} - \overset{63}{(-48-15)} = 35.$$

Therefore, $x = (1, -2, 1)$ is the unique solution. (Check that it actually solves the system!)

- Topology in $\mathbb{R}^n$ (See Simon + Blume, Ch 12)

- All the set theory from the beginning of the course applies when $X = \mathbb{R}^n$. We can define an order, denoted $\geq$, on $\mathbb{R}^n$ as

$$x \geq y \quad \text{iff} \quad \forall j = 1 \dots n \quad x_j \geq y_j.$$

When $n=1$, this is the usual greater-than-or-equal-to relation.

- In general, it is transitive, i.e., $x \geq y \geq z$ implies $x \geq z$. Unlike the greater-than-or-equal-relation on

$\mathbb{R}$, however, it is not complete. That is, it may be that neither $x \geq y$ nor $y \geq x$. To see this, take $(2,1) \in \mathbb{R}^2$ and $(1,2) \in \mathbb{R}^2$. Thus, we call $\geq$ a <u>partial order</u>.

- Upper and lower bounds are defined as before, on p. 41, where we let $Y \subseteq \mathbb{R}^n$ and x and y be vectors. Supremum and infimum may also be defined as before. Again, the supremum (and infimum) of a set is unique.

- A sequence is defined as before, on p. 47, but now we let x_1, x_2, etc. be vectors in $\mathbb{R}^n$, though we should write x^1, x^2, etc. We can't use n to index the sequence now, so let $\langle x^k \rangle$ denote a sequence in $\mathbb{R}^n$.

- <u>Convergence</u> is defined as before, but now use $\|\cdot\|$ instead of $|\cdot|$: $\forall \varepsilon > 0 \; \exists K$ s.t. $\forall k \geq K$ $\|x^k - x\| < \varepsilon$.

- We don't usually talk about "increasing" or "decreasing" sequences of vectors, though one could, I suppose. We don't talk about liminf or limsup in $\mathbb{R}^n$.

- The next theorem makes an important connection between convergence in $\mathbb{R}^n$ and convergence in $\mathbb{R}$.

PSC 403

- Theorem: The sequence $\langle x^k \rangle$ in $\mathbb{R}^n$ converges to x iff for all $i = 1, \dots, n$, $\langle x_i^k \rangle$ converges to x_i.

- Proof: First suppose $x^k \to x$ in $\mathbb{R}^n$, and pick any i, and pick any $\varepsilon > 0$. Since $x^k \to x$, there is some K such that $k \geq K$ implies $\| x^k - x \| < \varepsilon$, i.e.,

$$\sqrt{(x_1^k - x_1)^2 + (x_2^k - x_2)^2 + \cdots + (x_n^k - x_n)^2} < \varepsilon.$$

Since each term in the square root is non-negative, this implies $\sqrt{(x_i^k - x_i)^2} < \varepsilon$, or equivalently, $|x_i^k - x_i| < \varepsilon$. Therefore, $x_i^k \to x_i$.

Now suppose that, for all i, $x_i^k \to x_i$. To show that $x^k \to x$, take any $\varepsilon > 0$. Since $x_1^k \to x_1$, there is some K_1 such that $k \geq K_1$ implies $|x_1^k - x_1| < \varepsilon/\sqrt{n}$. Similarly, there is some K_2 such that $k \geq K_2$ implies $|x_2^k - x_2| < \varepsilon/\sqrt{n}$, and so on. Let $K = \max\{K_1, K_2, \dots, K_n\}$. I claim that $k \geq K$ implies $\| x^k - x \| < \varepsilon$. Note that

$$\begin{aligned} \| x^k - x \| &= \sqrt{(x_1^k - x_1)^2 + (x_2^k - x_2)^2 + \cdots + (x_n^k - x_n)^2} \\ &< \sqrt{\varepsilon^2/n + \varepsilon^2/n + \cdots + \varepsilon^2/n} \\ &= \varepsilon. \end{aligned}$$

//

PSC 403

- Proposition 4.8 (i) holds just as before: Let $x^k \to x$ and $y^k \to y$, and let $a, b \in \mathbb{R}$. Then $ax^k + by^k \to ax + by$.

- The proof uses the previous Theorem. If $x^k \to x$ and $y^k \to y$ then $x_1^k \to x_1$ and $y_1^k \to y_1$. By the original proposition, $ax_1^k + by_1^k \to ax_1 + by_1$. Similarly, $ax_2^k + by_2^k \to ax_2 + by_2$, and so on. Using the theorem again, we see that $ax^k + by^k \to ax + by$.

- Prop 4.10 still holds. B1 Sandwich theorem

- A Cauchy sequence is defined as before, on p. 56, but using $\|\cdot\|$. Then Proposition 5.17 holds.

- A closed set is defined exactly as on p. 57. A subset $X \subseteq \mathbb{R}$ is <u>open</u> if, for all $x \in X$ there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq X$, where

$$B_\varepsilon(x) = \{ y \in \mathbb{R}^n \mid \|x - y\| < \varepsilon \}$$

is the open ball of radius ε around x.

- The proposition on p. 57 still holds.

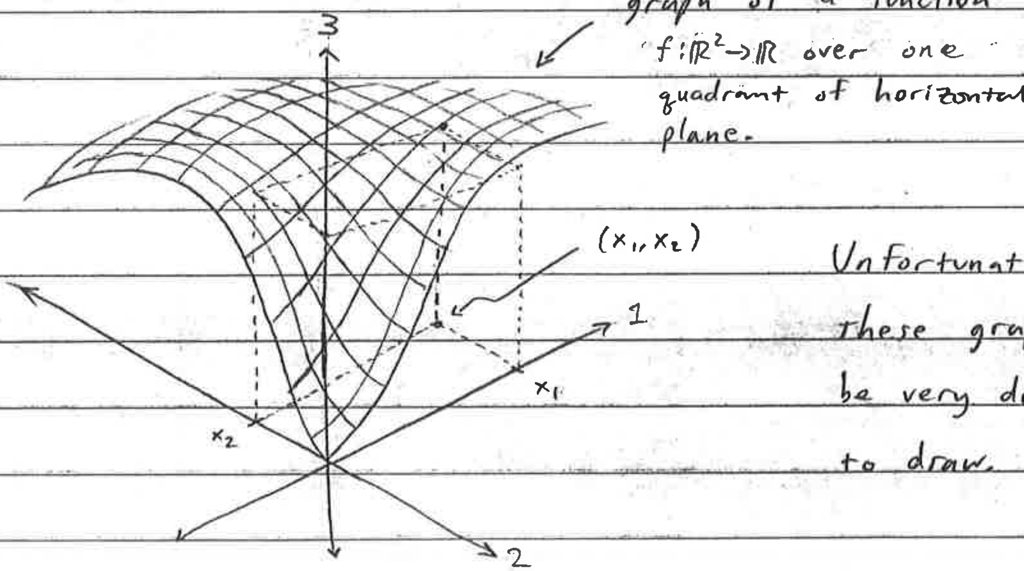
PSC 403

- A subsequence is defined just as before. Theorem 5.2, on p.59, still holds. A compact set is defined exactly as before.

- The Balzano-Weierstrauss Theorem, on p.60, continues to be true for all closed, bounded $X \subseteq \mathbb{R}^n$.

- Functions on $\mathbb{R}^n$

  - Like functions $f: \mathbb{R} \to \mathbb{R}$, functions $f: \mathbb{R}^2 \to \mathbb{R}$ can be graphed. Letting the horizontal plane represent the domain of f, the graph should indicate a vertical magnitude, $f(x_1, x_2)$, for each point (x_1, x_2) on the plane. The graph of f looks like the surface of a 3-dimensional object.



Unfortunately, these graphs can be very difficult to draw.

- A much simpler technique is to draw the "level sets" of f as subsets of $\mathbb{R}^2$.

- Given $f: X \to \mathbb{R}$, define the level set of f through x as

$$\{ y \in X \mid f(y) = f(x) \}.$$

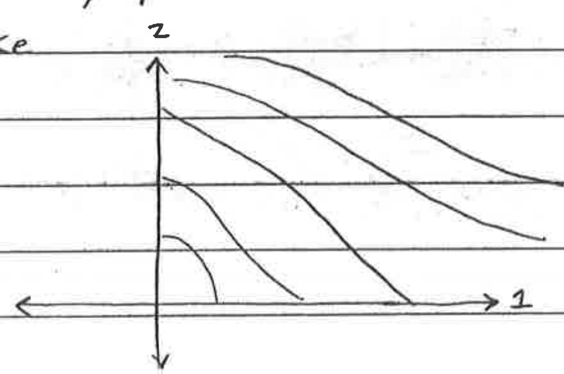
Given $a \in \mathbb{R}$, define the level set of f at a as

$$\{ x \in X \mid f(x) = a \}.$$

(Sometimes I might use the term contour set.)

- If $f: \mathbb{R}^2 \to \mathbb{R}$, the level sets of f will be subsets of $\mathbb{R}^2$. If we graph a few level sets of f and label them with the value taken by f on those sets, we can get a good idea of the structure of f.

- For the function I graphed above, the level sets might look like

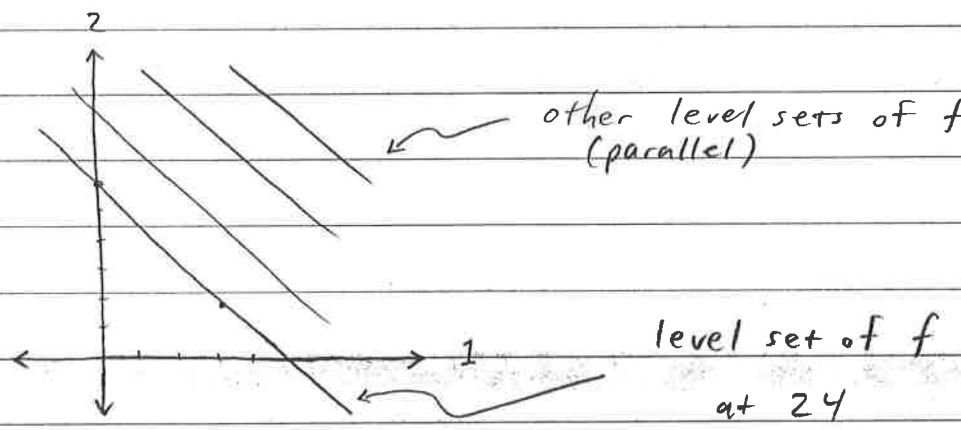


PS403

- Examples:

(i) $f(x_1, x_2) = 3x_1 + 4x_2$

To find the level set of f at a,
note that $f(x_1, x_2) = a$ iff
$3x_1 + 4x_2 = a$ iff $x_2 = \frac{1}{4}a - \frac{3}{4}x_1$.



(ii) $f(x_1, x_2) = -\sqrt{(x_1 - 3)^2 + (x_2 - 4)^2}$

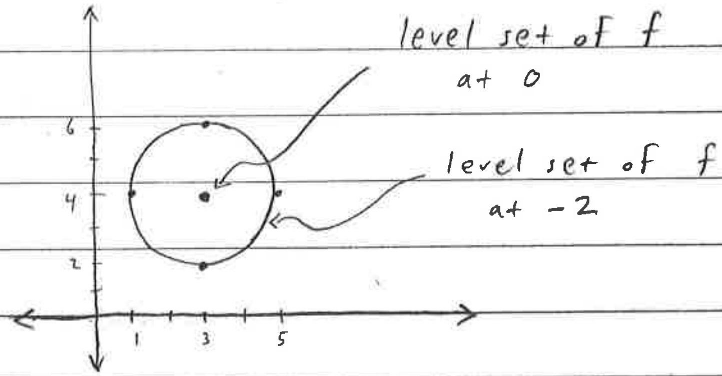
Note that $f(x_1, x_2) = a$ iff

$$(x_1 - 3)^2 + (x_2 - 4)^2 = a^2$$

iff $(x_2 - 4)^2 = a^2 - (x_1 - 3)^2$
iff $x_2 - 4 = \pm\sqrt{a^2 - (x_1 - 3)^2}$
iff $x_2 = 4 \pm \sqrt{a^2 - (x_1 - 3)^2}$

PSC 403

Graphing the level set of f at a, Spatial model



Note: It's easy to see that this function will have circular level sets if you write it as

$$f(x_1, x_2) = -\|(x_1, x_2) - (3,4)\|.$$

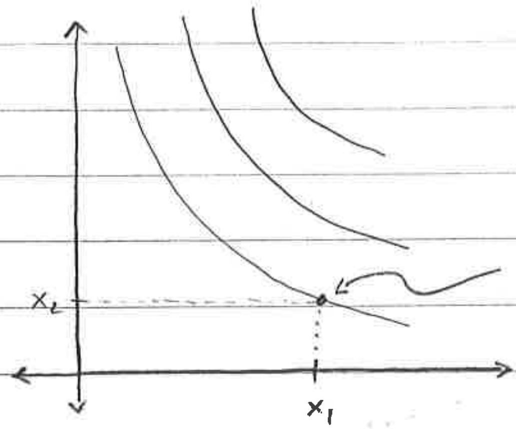
So, the function just says how far (x_1, x_2) is from $(3,4)$ (times -1). So the level set of f at -2 is just the set of points at distance 2 from $(3,4)$.

Note: This function has a maximizer at $(3,4)$.

(iii) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$, $\alpha \in (0,1)$.

Here, $f(x_1, x_2) = a$ iff $x_1^{\alpha} x_2^{1-\alpha} = a$
iff $x_2 = \left(a / x_1^{\alpha}\right)^{\frac{1}{1-\alpha}} = a^{\frac{1}{1-\alpha}} x_1^{-\frac{\alpha}{1-\alpha}}$.

PSC 403



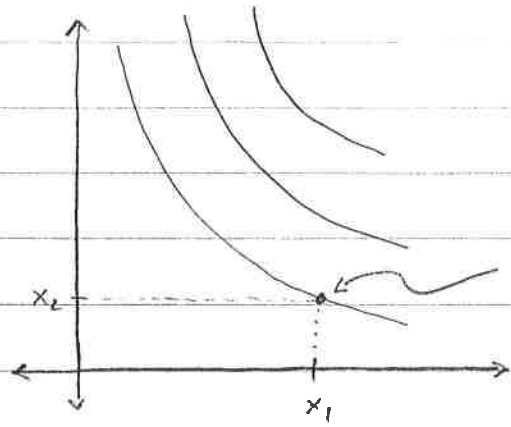
slope $= -\left(\frac{\alpha}{1-\alpha}\right) a^{\frac{1}{1-\alpha}} x_1^{-\frac{1}{1-\alpha}}$

Note: The slope of each level set is negative and increasing (decreasing in absolute value) as x_1 increases.

- Intuitively, the level sets of f are like a "contour map" of the graph of f.

- linear functions...

PSC 403



slope $= -\left(\frac{\alpha}{1-\alpha}\right) a^{\frac{1}{1-\alpha}} x_1^{-\frac{1}{1-\alpha}}$

Note: The slope of each level set is negative and increasing (decreasing in absolute value) as x_1 increases.

- Intuitively, the level sets of f are like a "contour map" of the graph of f.

- Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is used to represent the preferences of a voter or consumer, i.e., $f(x) > f(y)$ iff x is better than y. Then we call f a utility function. (For obvious reasons, we usually use "u" instead of "f.") Looking at the level sets of f gives us an idea of the consumer's preferences. The level sets of a utility function are called indifference curves or indifference contours.

- When $n=2$, $u(x_1, x_2) = \alpha x_1 + \beta x_2$, $u(x_1, x_2) = -\|(x_1, x_2) - (\alpha, \beta)\|$, and $u(x_1, x_2) = x_1^\alpha x_2^\beta$ are typical examples of utility functions. More generally, Fixing $\alpha_1, \alpha_2 \cdots \alpha_n \in \mathbb{R}$,

$$u(x_1, x_2 \cdots x_n) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$
(<u>linear preferences</u>)

$$u(x_1, x_2 \cdots x_n) = -\| (x_1, x_2 \cdots x_n) - (\alpha_1, \alpha_2 \cdots \alpha_n) \|$$
(<u>Euclidean preferences</u>)

$$u(x_1, x_2, \cdots x_n) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$
(<u>Cobb-Douglas preferences</u>)

- When $\alpha_1, \alpha_2 \cdots \alpha_n > 0$, linear and Cobb-Douglas preferences are often used to represent the preferences of consumers over consumption levels of n commodities — here, x_j is the level of consumption of the jth commodity. The vector $(x_1, x_2 \cdots x_n)$ is then called a <u>consumption bundle</u> or simply <u>bundle</u>.

- Euclidean preferences are used to represent preferences of voters when x_j represents the amount of a fixed resource (the federal budget, for example) allocated to some activity (national defense, for example). Or x_j may represent a more qualitative characteristic of policy outcomes: it may measure how liberal a policy is, or how moral, or how environmentally sound, etc. Here, the consumer's favorite policy, or <u>ideal point</u>, is $(\alpha_1, \alpha_2 \cdots \alpha_n)$.

PSC 403

- The model with Euclidean preferences is often referred to as the <u>spatial model</u>.

- A function $f: \mathbb{R}^n \to \mathbb{R}$ is <u>linear</u> if, for all $a, b \in \mathbb{R}$ and all $x, y \in \mathbb{R}^n$, $f(ax+by) = af(x) + bf(y)$.

- Breaking this into two parts, f is linear if

(i) $f(ax) = af(x)$

and

(ii) $f(x+y) = f(x) + f(y)$.

- Proposition: $f: \mathbb{R}^n \to \mathbb{R}$ is linear if and only if there exists $a = (a_1, a_2 \cdots a_n) \in \mathbb{R}^n$ such that, for all $x \in \mathbb{R}^n$,

$$f(x) = a \cdot x.$$

- Proof: I'll omit the $\Leftarrow$ direction. To prove $\Rightarrow$, consider any linear function f. For each unit coordinate vector e^j, define $a_j = f(e^j)$. I claim that $a = (a_1, a_2 \cdots a_n)$, so-defined, fulfills the proposition. Take any $x = (x_1, x_2 \cdots x_n) \in \mathbb{R}^n$, and note that

$$x = x_1 e^1 + x_2 e^2 + \cdots + x_n e^n.$$

Then

$$\begin{aligned} f(x) &= f(x_1 e^1 + x_2 e^2 + \cdots + x_n e^n) \\ &= x_1 f(e^1) + x_2 f(e^2) + \cdots + x_n f(e^n) \\ &= x_1 a_1 + x_2 a_2 + \cdots + x_n a_n \end{aligned}$$

PSC 403

$$= x \cdot a,$$

as claimed. //

– What do the contour sets of a linear function look like? When $n = 2$, suppose $f(x) = a \cdot x$, where $a = (a_1, a_2)$. Then the level set at zero is defined by the equation $f(x) = 0$, i.e.,

$$a_1 x_1 + a_2 x_2 = 0$$

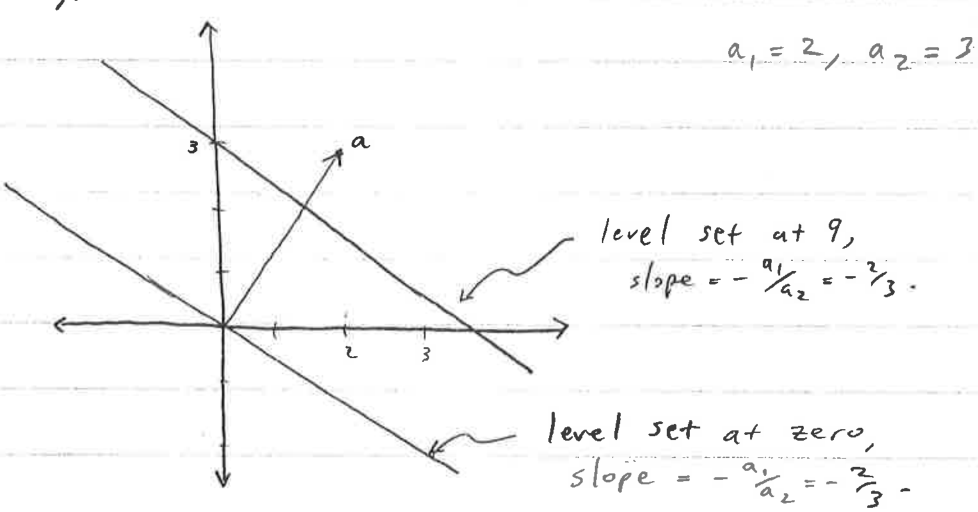
or

$$x_2 = -\frac{a_1}{a_2} x_1.$$

In general, the level set at c is given by

$$x_2 = \frac{c}{a_2} - \frac{a_1}{a_2} x_1.$$

Graphing,



PSC 403

- Note that the level set at zero is perpendicular to a. That's no surprise, because it consists of the vectors x such that $a \cdot x = 0$.

- In general, the level set at zero is the subspace (of dimension $n-1$) orthogonal to $a = (a_1, \dots a_n)$.

- Note that the level sets of a linear function $f : \mathbb{R}^2 \to \mathbb{R}$ are parallel — they are all perpendicular to $a = (a_1, \cdots a_n)$.

- In general, if Y denotes the level set at zero, then Z is a level set of f if and only if there is some $x \in \mathbb{R}^n$ such that

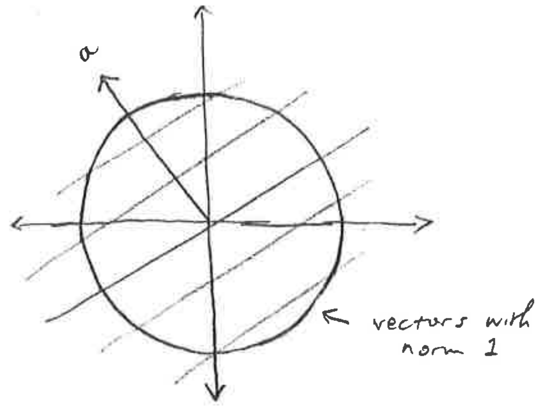
$$Z = \{ y + x \mid y \in Y \}.$$

(This set is sometimes denoted $Y + x$.)

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is <u>affine</u> if there is some $c \in \mathbb{R}$ such that $f(x) + a$ is linear.

- Take a linear function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = a \cdot x$, and think about this: What vector y maximizes f on the set $\{ z \in \mathbb{R}^n \mid \|z\| = 1 \}$?

PSC 403

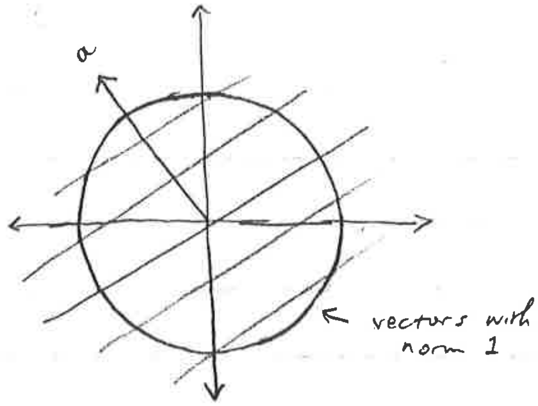


- The answer is $\frac{1}{\|a\|} a$. Later on, we will call a the <u>gradient</u> of f, and we have just seen: the gradient indicates the direction in which f is increasing the fastest.

- Why this answer? Given any $z \in \mathbb{R}^n$ with $\|z\|=1$, we know

$$f(z) = a \cdot z = \|a\| \, \|z\| \cos\alpha = \|a\| \cos\alpha .$$

This is maximized when $\alpha = 0$, or z points in the same direction as a.



- The answer is $\frac{1}{\|a\|} a$. Later on, we will call this the <u>gradient</u> of f, and we have just seen: the gradient indicates the direction in which f is increasing the fastest.

- Why this answer? Given any $z \in \mathbb{R}^n$ with $\|z\|=1$, we know

$$f(z) = a \cdot z = \|a\| \|z\| \cos\alpha = \|a\| \cos\alpha.$$

This is maximized when $\alpha = 0$, or z points in the same direction as a.

NOTE: THE GRAPH OF f IS A PLANE THROUGH THE ORIGIN. IT'S SLOPE IN THE DIRECTION OF THE jTH AXIS IS a_j. THEREFORE: THE SLOPE IN THE DIRECTION y, WHERE $\|y\|=1$, IS $a \cdot y$. (move y_1 units in x_1 direction, go up $a_1 y_1$. move y_2 in x_2 direction, go up $a_2 y_2$. Total up: $a_1 y_1 + a_2 y_2$. Total over: 1, since $\|y\|=1$.)

- A function $f: \mathbb{R}^n \to \mathbb{R}$ is <u>concave</u> if, for all $x, y \in \mathbb{R}^n$ and all $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha) y) \geq \alpha f(x) + (1-\alpha) f(y)$. It is <u>strictly concave</u> if this expression holds with a strict inequality for all $x \neq y$ and $\alpha \in (0,1)$. <u>Convexity</u> and <u>strict convexity</u> are also defined as on p. 64.

PSC 403

Related to the idea of a level set are the following concepts.

I skipped over this stuff in class. We'll come back to it when we need it.

weak upper contour set at a $\{x \in \mathbb{R}^n \mid f(x) \geq a\}$

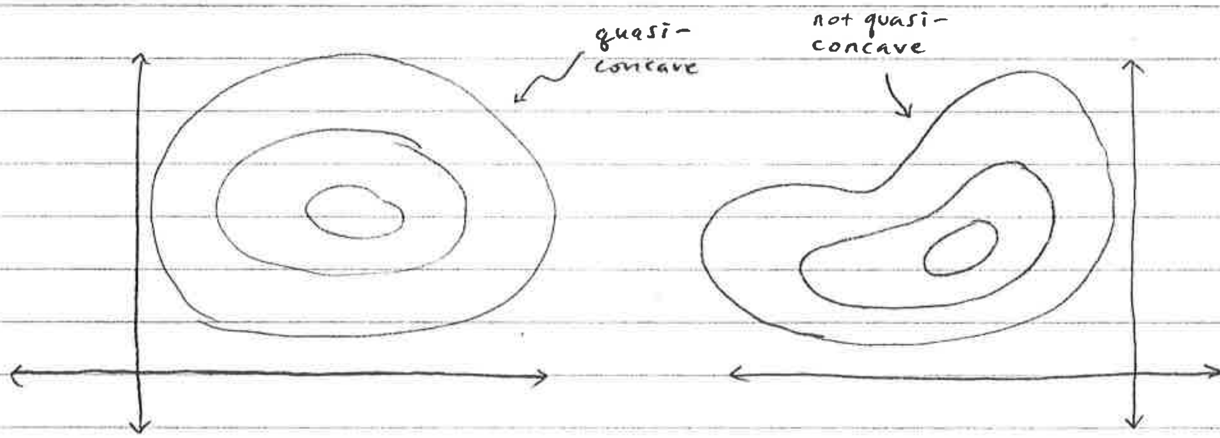
weak lower contour set at a $\{x \in \mathbb{R}^n \mid f(x) \leq a\}$.

- A function $f: \mathbb{R}^n \to \mathbb{R}$ is quasi-concave if, for all $x, y \in \mathbb{R}^n$ and all $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \geq \min\{f(x), f(y)\}$. It is quasi-convex if $f(\alpha x + (1-\alpha)y) \leq \max\{f(x), f(y)\}$.

- You can check that quasi-concavity is equivalent to "convex" weak upper contour sets, i.e., if $Y \subseteq \mathbb{R}^n$ is the weak upper contour set at a, and $\alpha \in [0,1]$

This definition of convex is the same as the one on p. 69.

$$x, y \in Y \Rightarrow \alpha x + (1-\alpha)y \in Y.$$



- The definitions of "bounded above," "bounded below," and "bounded" are just as on p. 65.

- The definition of a maximizer and minimizer is as on p.65.

- Given $X \subseteq \mathbb{R}^n$, $f: X \to \mathbb{R}$, and $x \in X$, f is continuous at x if, for every sequence $\langle x^k \rangle$ in X converging to x, $f(x^k) \to f(x)$. (Exactly as on p.66.) It is continuous if it is continuous at every $x \in X$.

closed weak upper and weak lower contour sets...

- Theorem 8.12 (combinations of continuous functions are continuous) and Theorem 8.17 (compositions of continuous functions are continuous) carry over to $f: \mathbb{R}^n \to \mathbb{R}$.

- So do Theorems 9.9 (Intermediate Value Theorem), 9.12, and the following corollary (continuous functions on compact sets have maximizers and minimizers).

- We say $\langle x^k \rangle$ eventually converges to x if $x^k \to x$ and, for all k, $x^k \neq x$. (As on p.71.)

- Given $X \subseteq \mathbb{R}^n$ and $x \in X$, $f: X \to \mathbb{R}$ is differentiable at x if there is a linear function $g: \mathbb{R}^n \to \mathbb{R}$ such that, for every sequence $\langle h^k \rangle$ eventually converging to 0,

$f: \mathbb{R} \to \mathbb{R}$

derivative as slope of graph

derivative as affine approximation

$$\frac{|f(x+h^k) - f(x) - g(h^k)|}{\|h^k\|} \to 0.$$

By our earlier result, this is equivalent to: there exists $a \in \mathbb{R}^n$ such that, for every sequence $\langle h^k \rangle$ eventually converging to 0,

$$\frac{|f(x+h^k) - f(x) - a \cdot h^k|}{\|h^k\|} \to 0.$$

PSC403

- In this definition, we can think of there being an affine function $G : \mathbb{R}^n \to \mathbb{R}$ defined by

$$G(y) = f(x) - g(x) + g(y)$$

where g is the linear function above. I claim the graph of G is tangent to the graph of f at x. Note that $G(x) = x$, and, for all $h^k \to 0$,

$$\frac{|f(x+h^k) - f(x) - g(h^k)|}{\|h^k\|}$$

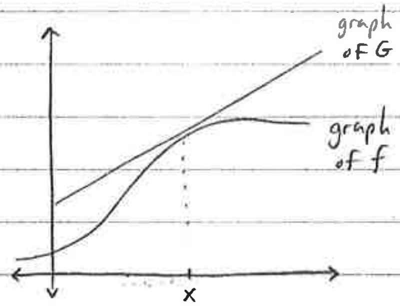
$$= \frac{|f(x+h^k) - f(x) + g(x) - g(x+h^k)|}{\|h^k\|}$$

$$= \frac{|f(x+h^k) - G(x+h^k)|}{\|h^k\|}$$

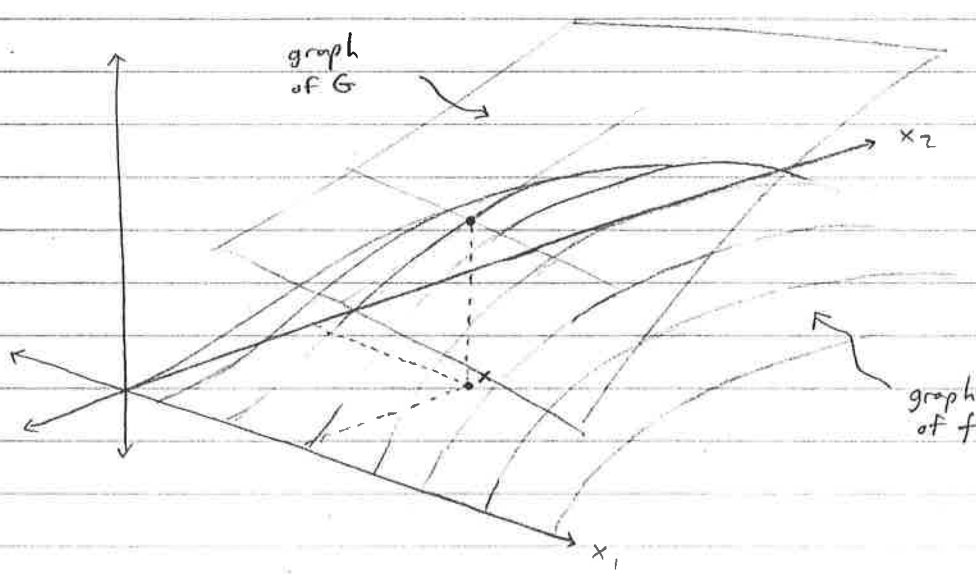
$$\to 0,$$

as claimed.

$\boxed{n=1}$



$\boxed{n=2}$



PSC 403

- The linear function $g: \mathbb{R}^n \to \mathbb{R}$ is called the derivative of f at x.

- We will soon see that the vector $a \in \mathbb{R}^n$ defining g has a familiar form.

- Recall that, for $f: \mathbb{R} \to \mathbb{R}$, we could think of the derivative as the slope of the graph of f. When $f: \mathbb{R}^2 \to \mathbb{R}$ or $f: \mathbb{R}^n \to \mathbb{R}$, of course, the "slope of the graph" will depend on the direction you're moving.

- Given $f: \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, and $t \in \mathbb{R}^n$ with $\|t\| = 1$, define the function $g: \mathbb{R} \to \mathbb{R}$ by

$$g(\alpha) = f(x + \alpha t).$$

We say f is differentiable at x in direction t if g is differentiable at 0. The directional derivative at x in direction t is

$$D_t f(x) = Dg(0).$$

- Intuitively, span$\{t\}$ is the line through 0 and t. We can think of f, restricted to this line, as a function of real numbers. The directional derivative is just the derivative of f, restricted to the line.

- Given a unit coordinate vector e^k, the derivative of f at x in the direction e^k, i.e., $D_{e^k} f(x)$, has a special

PSC 407

designation. We call it the k^{th} partial derivative of f at x, and we use the notation

$$\frac{\partial f}{\partial x_k}(x).$$

- The vector of partial derivatives at x,

$$\left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)$$

is called the gradient of f at x and is denoted by $\nabla f(x)$.

- Let's connect this with what we've learned about derivates. Let $a \in \mathbb{R}^n$ represent the derivate of f. As we've discussed, given $t \in \mathbb{R}^n$ with $\|t\| = 1$, the slope of the graph of f in the direction t is $a \cdot t$. This means,

$$D_t f(x) = a \cdot t,$$

where a is the derivative of f.

- Setting $t = e^k$,

$$\frac{\partial f}{\partial x_k}(x) = D_{e^k} f(x) = a \cdot e^k = a_k.$$

So the partial derivatives of f give us the derivative! In the terminology above the derivative of f is represented by the gradient:

$$a = \nabla f(x).$$

PSC 403

- Then the derivative in direction t is

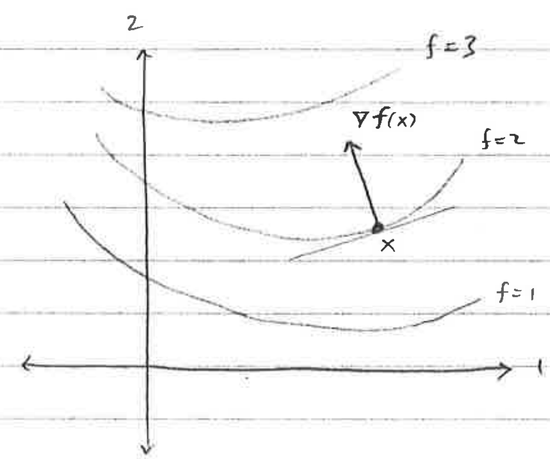
$$D_t f(x) = a \cdot t = \nabla f(x) \cdot t = \sum_{k=1}^{n} t_k \frac{\partial f}{\partial x_k}(x).$$

- At a point on the graph of f, what is the direction of steepest ascent? In other words, which t with $\|t\| = 1$ maximizes $D_t f(x)$? We've just learned that this is $\nabla f(x) \cdot t$. But recall our analysis of linear functions: If $g(x)$ is a linear function defined by $g(x) = a \cdot x$, the vector y with $\|y\| = 1$ maximizing $a \cdot y$ is just $\frac{1}{\|a\|} a$. So, what t with $\|t\| = 1$ maximizes $\nabla f(x) \cdot t$? It's just

$$\frac{1}{\|\nabla f(x)\|} \nabla f(x).$$

- In other words, the gradient of f points in the direction of steepest ascent up the graph of f.

- If $f: \mathbb{R}^2 \to \mathbb{R}$ has level sets as below, the gradient of f at x will look something like that pictured. If you draw a line tangent to the level set at x, $\nabla f(x)$ will be perpendicular to that line. (To prove this for yourself, you would take a

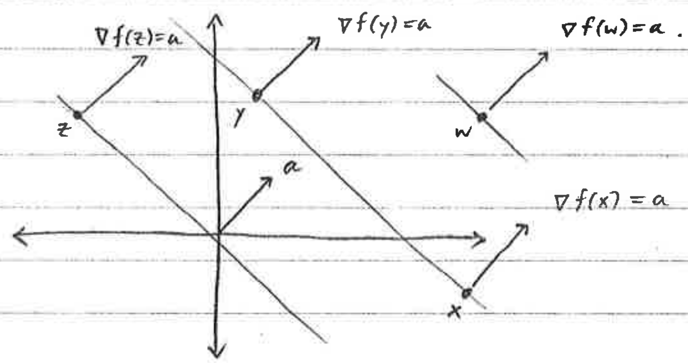


sequence $\langle x^k \rangle$ of points on the level set converging to x and

PSC 403

show that $\nabla f(x) \cdot (x^k - x) \to 0.$)

- If f is a linear function defined by $f(x) = a \cdot x$, then, for all x, $Df(x) = a$. (To see this, just check the definition of the derivative.) Therefore $\nabla f(x)$ is independent of x.



- A nice property of partial derivatives is that they are easy to calculate. Going back to the definition of the directional derivative,

$$\frac{\partial f}{\partial x_j}(x) = D_{e^j} f(x) = D g(0),$$

where

$$g(\alpha) = f(x + \alpha e^j).$$

Thus, taking a sequence $\langle h_k \rangle$ of real numbers with $h_k \to 0$,

$$\frac{\partial f}{\partial x_j}(x) = \lim_{k \to \infty} \frac{f(x_1, \ldots x_{j-1}, x_j + h_k, x_{j+1} \cdots x_n) - f(x)}{h_k}$$

It follows that we can fix $x_1 \ldots x_{j-1}, x_{j+1} \ldots x_n$, think of f as a function of x_j alone, and just take the usual derivative with respect to this variable.

PSC 403

– Examples: If $n=2$ and, for fixed $y \in \mathbb{R}^2$,

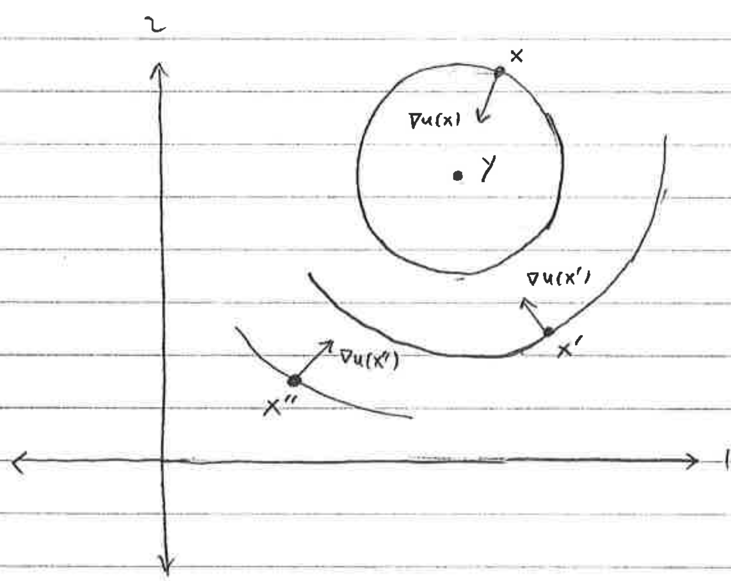
$$u(x) = -\|x-y\| = -\sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$$

Then

$$\frac{\partial u}{\partial x_1}(x) = -\frac{1}{2}\left((x_1-y_1)^2 + (x_2-y_2)^2 \right)^{-\frac{1}{2}} 2(x_1-y_1)$$

$$= -\frac{x_1-y_1}{\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}} = -\frac{x_1-y_1}{\|x-y\|}$$

$$\frac{\partial u}{\partial x_2}(x) = -\frac{x_2-y_2}{\|x-y\|}$$



Note that $\nabla u(x) = \frac{1}{\|x-y\|}(y_1-x_1, y_2-x_2) = \frac{1}{\|x-y\|}(y-x)$.

In other words, $\nabla u(x)$ is the unit vector pointing from x to y.

PSC 403

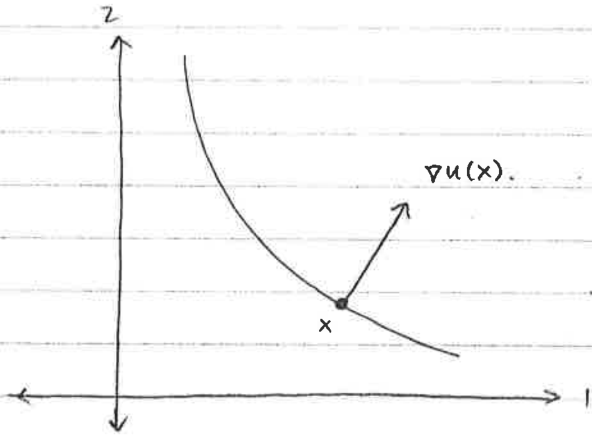
- If $n=2$ and $u(x) = x_1^{\alpha} x_2^{1-\alpha}$, for fixed $\alpha \in [0,1]$, then

$$\frac{\partial u}{\partial x_1}(x) = \alpha x_1^{\alpha-1} x_2^{1-\alpha} = \alpha \left(\frac{x_2}{x_1}\right)^{1-\alpha}$$

$$\frac{\partial u}{\partial x_2}(x) = (1-\alpha)\left(\frac{x_1}{x_2}\right)^{\alpha}.$$

Therefore,

$$\nabla u(x) = \left(\alpha \left(\frac{x_2}{x_1}\right)^{1-\alpha}, (1-\alpha)\left(\frac{x_1}{x_2}\right)^{\alpha} \right).$$

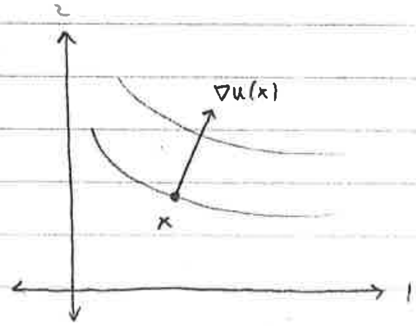


Note the "slope" of gradient vector: it is

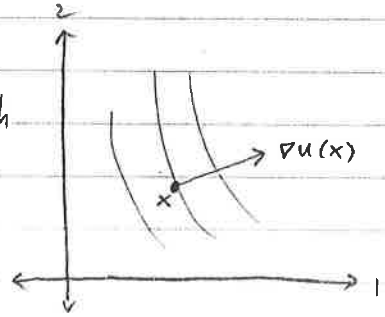
$$\frac{\frac{\partial u}{\partial x_2}(x)}{\frac{\partial u}{\partial x_1}(x)} = \frac{(1-\alpha)\left(\frac{x_1}{x_2}\right)^{\alpha}}{\alpha\left(\frac{x_2}{x_1}\right)^{1-\alpha}} = \frac{(1-\alpha)}{\alpha}\left(\frac{x_1}{x_2}\right).$$

If α increases, so the utility function puts more weight on good 1, the gradient points more to the east.

low α

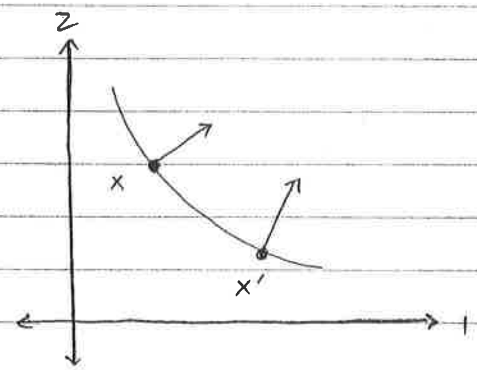


high α



This is because, as α gets bigger, the "utility hill" transforms shape, becoming steeper in the direction of the x_1 axis.

If x_1/x_2 increases, the gradient points more toward the north. If the consumer is consuming a lot of good 1 relative to good 2 and wants to increase his utility, the best way to do that won't be to consume a lot more good 1 relative to good 2 — he already has an abundance of good 1.



- A very important concept in rational choice theory is that of marginal rate of substitution. The <u>marginal rate of substitution of good 1 with respect to good 2 at x</u> is defined, when $\partial u/\partial x_2(x) \neq 0$, as

$$MRS_{1,2}(x) = \frac{\frac{\partial u}{\partial x_1}(x)}{\frac{\partial u}{\partial x_2}(x)}.$$

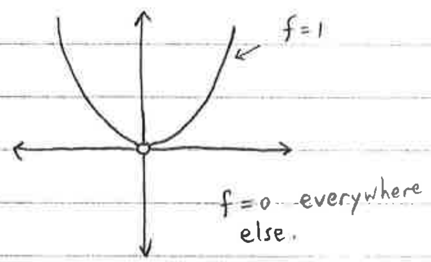
- Notice, this is one over the "slope" of the gradient vector. Technically, it is negative one times the slope of the indifference curve through x at x. (I'll come back to this.)

- Intuitively, the $MRS_{1,2}(x)$ measures the value of a small increase in the amount of good 1 consumed relative to the value of a small increase in the amount of good 2 consumed.

PSC 403

- We've seen that if a function is differentiable at x then it has directional derivatives at x in every direction. But the converse does not necessarily hold.

- Consider $f: \mathbb{R}^2 \to \mathbb{R}$ defined below.



$f(x)=1$ on the indicated parabola, but not at the origin. Everywhere else, $f(x)=0$. Here, f has directional derivatives at 0 in every direction. Note: we could "smooth" f out along the parabola.

- Proposition: Let $f: \mathbb{R}^n \to \mathbb{R}$ have partial derivatives with respect to each unit coordinate vector everywhere. If at every point $\frac{\partial f}{\partial x_1}(x), \ldots \frac{\partial f}{\partial x_n}(x)$ are all continuous at x then f is differentiable at x.

see Binmore

- Proposition: Let $f: \mathbb{R}^n \to \mathbb{R}$, $g_1: \mathbb{R} \to \mathbb{R}$, $g_2: \mathbb{R} \to \mathbb{R}, \ldots, g_n: \mathbb{R} \to \mathbb{R}$ be differentiable at x. Define $h: \mathbb{R} \to \mathbb{R}$ by

see Simon & Blume Thm 14.1.

$$h(x) = f(g_1(x), \ldots, g_n(x)).$$

Then

$$Dh(x) = \nabla f(g_1(x) \ldots g_n(x)) \cdot (Dg_1(x), \ldots, Dg_n(x))$$

$$= \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(g_1(x) \ldots g_n(x)) \frac{dg_j}{dx}(x).$$

- This version of the Chain Rule can be used to verify the connection between derivatives and directional derivatives. Recall, $D_t f(x) = \nabla f(x) \cdot t$. The definition of $D_t f(x)$ was $Dg(0)$, where

$$g(\alpha) = f(x + \alpha t) = f(x_1 + \alpha t_1, x_2 + \alpha t_2, \ldots, x_n + \alpha t_n).$$

PSC 403

Applying the chain rule with $g_j(\alpha) = x_j + \alpha t_j$, we see

$$Dg(\alpha) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x + \alpha t)\, t_j .$$

And when $\alpha = 0$, $Dg(0) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x)\, t_j = \nabla f(x) \cdot t$.

- If f has differentiable partial derivatives, i.e., $\frac{\partial f}{\partial x_j}(x)$ is differentiable as a function of x, we can define derivatives of higher orders. The i,j cross partial of f at x is

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x).$$

When $i = j$, we write $\frac{\partial^2 f}{\partial x_i^2}(x)$ for this.

- The Hessian of f at x is the matrix of cross partial derivatives,

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

- Observe the notation used for the Hessian: it looks like a "second derivative." We know the first derivative $Df(x)$ is a vector, $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x) \cdots \frac{\partial f}{\partial x_n}(x)\right)$. What would it mean to take the "derivative" of a "vector-function"? In fact, though we won't have time to go into it in detail, such a thing can be defined. (See Simon + Blume, Section 14.7.) It turns out that the Hessian is, indeed, the derivative of $\nabla f(x)$.

- Theorem ("Young's Theorem," Simon + Blume Thm 14.5):
  If $f : \mathbb{R}^n \to \mathbb{R}$ has continuous cross partials, i.e., $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ is continuous at x for all i, j, and x, then

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

  for all i, j, and x.

- In words, the order in which we take partial derivatives in calculating a cross partial is immaterial. In other words, $D^2 f(x)$ is a symmetric matrix.

- The next theorem is a simple version of the Implicit Function Theorem for the case $n=2$. It follows Simon & Blume's Theorem 15.1.

- Theorem: Let $f : \mathbb{R}^2 \to \mathbb{R}$ have cross partials continuous at $(\hat{x}, \hat{y})$. Suppose $f(\hat{x}, \hat{y}) = c$. If $\frac{\partial f}{\partial y}(\hat{x}, \hat{y}) \neq 0$ then there exists an open interval I containing $\hat{x}$ and a function $g : I \to \mathbb{R}$ such that

  (i) $g(\hat{x}) = \hat{y}$
  (ii) for all $x \in I$, $f(x, g(x)) = c$
  (iii) g is differentiable at $\hat{x}$
  (iv)
$$Dg(\hat{x}) = -\frac{\partial f/\partial x_1 (\hat{x}, \hat{y})}{\partial f/\partial x_2 (\hat{x}, \hat{y})}.$$

- Partial proof: Given (i), (ii), and (iii), we can derive (iv) using the chain rule. For all $x \in I$, $f(x, g(x)) = c$. Differentiating both sides at $\hat{x}$, we get

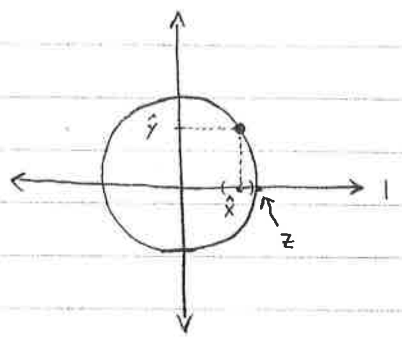
PSC 403

$$\frac{\partial f}{\partial x_1}(\hat{x}, g(\hat{x})) + \frac{\partial f}{\partial x_2}(\hat{x}, g(\hat{x}))\, Dg(\hat{x}) = 0,$$

or

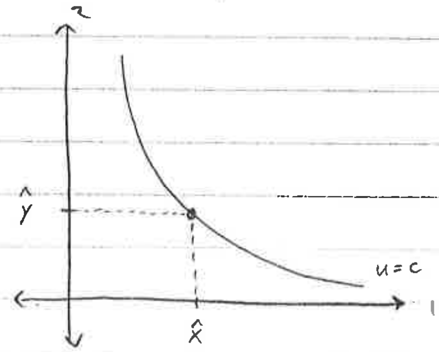
$$Dg(\hat{x}) = -\frac{\partial f/\partial x_1 (\hat{x},\hat{y})}{\partial f/\partial x_2 (\hat{x},\hat{y})}. \qquad //$$

- The typical example is $f(x,y) = x^2+y^2$. Fixing c, the level set of f at c is a circle. Thus, the graph of this level set does not determine a functional relationship between x and y. Indeed, at $\hat{x}$ there are two values of y, namely $\hat{y}$ and $-\hat{y}$, that yield $f(\hat{x},y)=c$. So let's pick one of those values. Moreover, let's pick one value of y for every x around $\hat{x}$ such that $f(x,y)=c$. The Implicit Function Theorem says we can do this in a differentiable way, and it tells us what the derivative of y is as a function of x.



But at z we can't do this: for x's to the right of z, there are no y's such that $f(x,y)=c$. The problem is that $\frac{\partial f}{\partial x_2}(z,0)=0$.

- I used the Implicit Function Theorem when I referred to the slope of a consumer's indifference curve. Consider the level set of a utility function u at c. Pick any $(\hat{x},\hat{y})$ on the level set. The theorem tells us that, under certain conditions, the slope of the indifference curve at



PSC 403

$(\hat{x}, \hat{y})$ is

$$-\frac{\partial u / \partial x_1 \, (\hat{x}, \hat{y})}{\partial u / \partial x_2 \, (\hat{x}, \hat{y})},$$

as I claimed earlier.

- Optimization in $\mathbb{R}^n$

- We can use the Hessian to calculate the second directional derivative of f at x in a given direction t. Of course, the first directional derivative is ...

$$D_t f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, t_i$$

- Fixing t, we can think of this as a function of x. We could call it $h(x) = D_t f(x) \cdot t$. As such, we can take partial derivatives:

$$\frac{\partial h}{\partial x_j}(x) = \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, t_i \right)$$

$$= \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x)\, t_i \, .$$

- We want to know how the directional derivative of f changes as we move in the direction t. That is, we want to know $D_t h(x)$. As always,

$$D_t h(x) = \sum_{j=1}^{n} \frac{\partial h}{\partial x_j}(x)\, t_j$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x)\, t_i \right) t_j$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x)\, t_i t_j \, .$$

- You can check that this is equal to

$$\underset{(1\times n)}{(t_1 \cdots t_n)} \underset{(n\times n)}{\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}} \underset{(n\times 1)}{\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}} .$$

- That is, the second directional derivative, denoted $D_t^2 f(x)$, is

$$D_t^2 f(x) = t' D^2 f(x) t,$$

where $t' = (t_1 \ldots t_n)$ and $t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$. Thus, the Hessian of f has an important use.

- Example: Let $n=2$, $f(x_1, x_2) = x_1 x_2 + x_1^3$. Then

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = x_2 + 3x_1^2 \qquad \frac{\partial f}{\partial x_2}(x_1, x_2) = x_1$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) = 1 \qquad \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) = 1$$

$$\frac{\partial^2 f}{\partial x_1^2}(x_1, x_2) = 6x_1 \qquad \frac{\partial^2 f}{\partial x_2^2} = 0.$$

Thus

$$D^2 f(x) = \begin{pmatrix} 6x_1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The derivative in direction $t = (t_1, t_2)$ is

$$D_t f(x) = (x_2 + 3x_1^2) t_1 + x_1 t_2$$

and the second derivative in direction t is

$$D_t^2 f(x) = (t_1 \; t_2) \begin{pmatrix} 6x_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

$$= (t_1 \; t_2) \begin{pmatrix} 6xt_1 + t_2 \\ t_1 \end{pmatrix}$$

$$= 6x_1 t_1^2 + 2t_1 t_2.$$

PSC 403

$(\hat{x},\hat{y})$ is
$$-\frac{\partial u/\partial x_1\,(\hat{x},\hat{y})}{\partial u/\partial x_2\,(\hat{x},\hat{y})},$$
as I claimed earlier.

- Optimization in $\mathbb{R}^n$

- We are often given a set $X \subseteq \mathbb{R}^n$ and $f: X \to \mathbb{R}$ and we want to find a maximizer of f. We want to solve
$$\max_{x \in X} f(x).$$
(Look familiar?)

- If X is compact and f is continuous, we know there is at least one maximizer.

- As before, x is a local maximizer of f if there is some $\varepsilon > 0$ such that, for all $y \in B_\varepsilon(x)$, $f(x) \geq f(y)$. It is a local minimizer if blah blah. We say x is a stationary point of f if $Df(x) = 0$.

- Theorem: Let $X \subseteq \mathbb{R}^n$, let x be in the interior of X, and let $f: X \to \mathbb{R}$ be differentiable at x. If x is a local maximizer or minimizer then $Df(x) = 0$.

- Proof: Take any unit coordinate vector e^i, and consider the restriction of f to the line through x in the direction e^i, essentially a function of real numbers only. Since x is a local maximizer of f in X, it is a local maximizer of (the restriction of) f to this line.

By our earlier theorem, the derivative of f at x along this line is zero, i.e., $\frac{\partial f}{\partial x_i}(x) = 0$. Since i was arbitrary,

$$Df(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) = 0. \qquad //$$

- Theorem: Let $X \subseteq \mathbb{R}^n$, let $f: X \to \mathbb{R}$ be twice differentiable on the interior of X, and let x be an interior point.

  (i) If x is a local maximizer of f then, for all $t \in \mathbb{R}^n$, $D_t^2 f(x) \leq 0$.

  (ii) If x is a local minimizer of f then, for all $t \in \mathbb{R}^n$, $D_t^2 f(x) \geq 0$.

- The proof is nearly identical to that of the previous theorem, using our earlier result for functions of one variable.

- As we've discussed, (i) can be replaced by $t' D^2 f(x) t \leq 0$ for all $t \in \mathbb{R}^n$. Some terminology: an $n \times n$ matrix A such that $x'Ax \leq$ for all $x \in \mathbb{R}^n$ is called <u>negative semi-definite</u>. Thus, (i) says that if x is a local maximizer of f then the Hessian of f at x is negative semi-definite.

- An $n \times n$ matrix A such that $x'Ax \geq 0$ for all $x \in \mathbb{R}^n$ is called <u>positive semi-definite</u>. Thus, if x is a local minimizer of f then the Hessian of f at x is positive semi-definite.

- We'll go over methods for checking semi-definiteness later.

PSC 403

- The two previous theorems give us necessary first and second order conditions for a local minimum or maximum. They are not, however, sufficient. As before, we need to strengthen the second order condition.

- Theorem: Let $X \subseteq \mathbb{R}^n$, let $f: X \to \mathbb{R}$ be twice differentiable on the interior of X, and let x be an interior point.

  (i) If $Df(x) = 0$ and $D_t^2 f(x) < 0$ for all $t \in \mathbb{R}^n$, then x is a local maximizer of f.

  (ii) If $Df(x) = 0$ and $D_t^2 f(x) > 0$ for all $t \in \mathbb{R}^n$, then x is a local minimizer of f.

- I won't go into the proof of this result. (See Binmore's Theorem 19.36.)

- An $n \times n$ matrix A such that $x'Ax < 0$ for all $x \in \mathbb{R}^n$ is called <u>negative definite</u>. If $x'Ax > 0$, it's <u>positive definite</u>.

- There is an interesting (and as far as I know, completely unintuitive) method for determining "definiteness" using determinants. The following discussion is taken from Simon + Blume's Ch. 16.

- Given an $n \times n$ matrix A, a $m \times m$ matrix formed by deleting $n-m$ rows of A and <u>the same</u> $n-m$ columns of A is called a <u>m^{th} order principal submatrix</u> of A. The determinant of the submatrix is a <u>m^{th} order principal minor</u> of A.

- The m^{th} order leading principal submatrix of A is that obtained by deleting the last $n-m$ rows and columns of A. The determinant of that submatrix is the m^{th} order leading principal minor of A. Let A_m denote the m^{th} order leading principal submatrix of A.

- Theorem (Simon + Blume, Thm 16.1) Let A be an $n \times n$ symmetric matrix.

  (i) A is positive definite iff all its n leading principal minors are strictly positive.

  (ii) A is negative definite iff its n leading principal minors alternate in sign as follows:

  $$|A_1| < 0, \quad |A_2| > 0, \quad |A_3| < 0 \ldots$$

  (iii) A is positive semi-definite iff every principal minor of A is non-negative.

  (iv) A is negative semi-definite iff every principal minor of odd order is non-positive and and every principal minor of even order is non-negative.

- Note: The conditions characterizing semi-definiteness are not nearly as convenient as those characterizing definiteness: we have to check the signs of all principal minors, not just the leading ones.

- Example: Consider a firm using two inputs, x_1 and x_2, in the production of some good. Suppose the "production plan" (x_1, x_2) costs $C(x_1, x_2) = x_1^2 + 2x_1x_2 + 4x_2^2$; the firm may sell as many units of output as it wants at price $p>0$; the most it can produce given (x_1, x_2) is output $F(x_1, x_2) = x_1 + x_1x_2$. The firm's problem is

$$\max_{x_1, x_2} \; p F(x_1, x_2) - C(x_1, x_2),$$

with the restriction, of course, that $x_1, x_2 \geq 0$. The necessary first order conditions are

$$p \frac{\partial F}{\partial x_1}(x_1, x_2) - \frac{\partial C}{\partial x_1}(x_1, x_2) = 0$$

$$p \frac{\partial F}{\partial x_2}(x_1, x_2) - \frac{\partial C}{\partial x_2}(x_1, x_2) = 0$$

or

$$p(1 + x_2) - (2x_1 + 2x_2) = 0$$

$$p x_1 - (2x_1 + 8x_2) = 0$$

or

$$-2x_1 + (p-2)x_2 = -p$$

$$(p-2)x_1 - 8x_2 = 0.$$

Since $\begin{pmatrix} -2 \\ p-2 \end{pmatrix}$ and $\begin{pmatrix} p-2 \\ -8 \end{pmatrix}$ are linearly independent for most prices (but not when $p=6$), there is a unique solution to the first order condition. Thus, there is a single candidate for an interior maximizer.

PSC 403

But remember that a stationary point can be a local minimum rather than a maximum, or neither. Let's verify the sufficient second order condition.

$$p\frac{\partial^2 F}{\partial x_1^2}(x_1,x_2) - \frac{\partial^2 C}{\partial x_1^2}(x_1,x_2) = -2$$

$$p\frac{\partial^2 F}{\partial x_2^2}(x_1,x_2) - \frac{\partial^2 C}{\partial x_2^2}(x_1,x_2) = -8$$

$$p\frac{\partial^2 F}{\partial x_1 \partial x_2}(x_1,x_2) - \frac{\partial^2 C}{\partial x_1 \partial x_2}(x_1,x_2) = p-2$$

The Hessian is actually independent of x_1 and x_2:

$$\begin{pmatrix} -2 & p-2 \\ p-2 & -8 \end{pmatrix}.$$

The leading principal minors are

$$\begin{vmatrix} -2 & p-2 \\ p-2 & -8 \end{vmatrix} = 16-(p-2)^2 \qquad \text{(2nd order)}$$

$$|-8| = -8. \qquad \text{(1st order)}$$

Thus $|A_1|<0$, as needed for a local maximizer. But we also need

$$|A_2| = 16-(p-2)^2 > 0$$

$$16 > (p-2)^2$$
$$4 > p-2$$
$$6 > p.$$

Thus, as long as the price of the firm's output is not too high (<6), we know that the solution to the firm's first order conditions is a local maximimizer.

Well, I'm getting ahead of myself. Let's use Cramer's rule to solve the first order conditions. We get

$$x_1^* = \frac{\begin{vmatrix} -p & p-2 \\ 0 & -8 \end{vmatrix}}{\begin{vmatrix} -2 & p-2 \\ p-2 & -8 \end{vmatrix}} = \frac{8p}{16-(p-2)^2}$$

$$x_2^* = \frac{\begin{vmatrix} -2 & -p \\ p-2 & 0 \end{vmatrix}}{\begin{vmatrix} -2 & p-2 \\ p-2 & -8 \end{vmatrix}} = \frac{p(p-2)}{16-(p-2)^2}.$$

Notice that, when $p=6$ there is no solution. When $p>6$, $x_1^*<0$. And when $p<2$, $x_2^*<0$. Clearly, optimization is not as simple as looking at first order conditions and solving!

Let's consider the cases:

$2<p<6$: (x_1^*, x_2^*) is a local maximizer. We will soon see that, in fact, it is a global maximizer.

$p \geq 6$: No local maximizer. The firm's profits are

$$\begin{aligned} pF(x_1,x_2) - C(x_1,x_2) &\geq 6F(x_1,x_2) - C(x_1,x_2) \\ &= 6x_1 + 4x_1x_2 - x_1^2 - 4x_2^2 \\ &= 6x_1 + (2x_1 - x_1)(2x_2 - x_2) \end{aligned}$$

$\geq 6x_1.$

Therefore, the firm can attain unlimited profits by choosing x_1 high enough. There is no global maximizer.

$p < 2$: No local maximizer. In this case, the firm's profits are

$$px_1 + px_1x_2 - x_1^2 - 2x_1x_2 - 4x_2^2$$

$$= px_1 + (p-2)x_1x_2 - x_1^2 - 4x_2^2.$$

Thus, as long as the firm uses a positive amount of x_1 (it will), it should set $x_2^* = 0$. Thus, the firm is essentially choosing x_1 to maximize its profits, while setting $x_2^* = 0$. You can check that the solution is $x_1^* = p/2$. The firm's maximized profits are then

$$px_1^* - x_1^{*2} = \frac{p^2}{2} - \frac{p^2}{4} = p^2/4.$$

$p = 2$: The firm's global maximizer is still $x_1^* = p^2/2$ and $x_2^* = 0$.

- NOTE ON BOTTOM OF P. 179: The last line should be

$$= 6x_1 - (x_1 - 2x_2)(x_1 - 2x_2).$$

So the firm can attain arbitrarily high profits by setting x_1 high and $x_2 = \frac{1}{2}x_1$. There is no global maximizer.

- So the example demonstrates two important things that can go wrong with the calculus approach to optimization:

① There may be no global maximizer! This was the case when $p \geq 6$. Although it wasn't the case in our example, you could actually have some local maxima, none of which are global maxizers. Be careful!

② There may be a global maximizer that's not an interior point! This was the case when $p \leq 2$. These maximizers can be tough to find, because they may not be critical points, as in the example. Again, there may be local maximizers that aren't global maximizers.

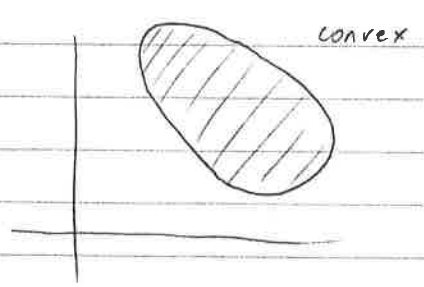
- What's to be done about these problems? ① If you don't know there's a global maximizer, you may very well be stuck. (Things are different for concave functions. See below.) In regard to this, remember: continuous functions on compact sets have maximizers! This was no help in the example, however, because our firm could choose any production plan in the non-negative orthant — not a compact set.

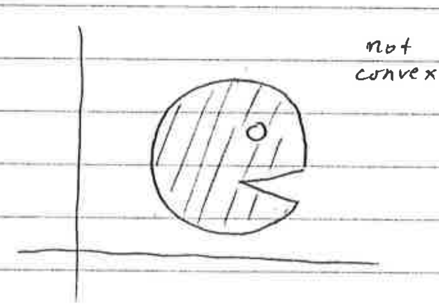
PSC 403

- ② The only thing to do is check the boundaries of the set you're maximizing over. You may have a local maximizer, but you won't know whether it's a global maximizer until you compare it with all other critical points and the boundary points. (Things are different for concave functions. See below.)

- To motivate our study of concave functions, go back to the $2<p<6$ case in the example. There we found a unique critical point satisfying the sufficient condition for a local maximizer. Though we cannot generally jump to the conclusion that we have a global maximizer, I claimed that we could in the example at hand. That's because, for prices $2<p<6$, the firm's profit function is concave.

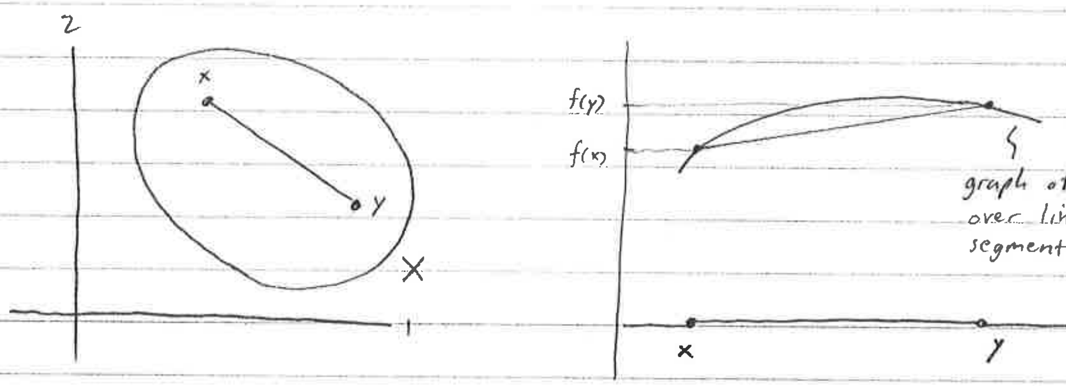
- A set $X \subseteq \mathbb{R}^n$ is convex if, for all $x, y \in X$ and all $\alpha \in [0,1]$, $\alpha x + (1-\alpha) y \in X$. That is, it's convex if the line segment connecting any two elements of X is itself included in X.





- Given convex $X \subseteq \mathbb{R}^n$, $f: X \to \mathbb{R}$ is concave if, for all $x, y \in X$ and all $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y)$. It is strictly concave if this expression holds with a strict inequality for all $x \neq y$ and $\alpha \in (0,1)$. It is convex if, for all $x, y \in X$ and all $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$. It is strictly convex if blah blah.

- Graphically, if we take any $x, y \in X$ and look at the



graph of f over the line segment connecting x and y, it must be on or above the line segment connecting $(x, f(x))$ and $(y, f(y))$. That is, it must be concave in the sense we've already seen for functions of one variable.

- The next theorem parallels our results for functions of one variable.

- Theorem: Let $X \subseteq \mathbb{R}^n$ be convex, let $f: X \to \mathbb{R}$ be differentiable on the interior of X, and let x be in the interior of X.

  (i) If f is concave and $Df(x) = 0$ then x is a global maximizer of f.

(ii) If f is convex and $Df(x)=0$ then x is a global minimizer of f.

- Proof: Consider (i). Let $Df(x)=0$ and suppose x is not a global maximizer. Then there is some $y \in X$ with $f(y) > f(x)$. Now think about the graph of f over the line through x and y. This is essentially the graph of a function of one variable, itself concave. Its derivative at x is the directional derivative of f in the $(x-y)$ direction, i.e., zero. From our earlier theorem for functions of one variable, x is a global maximizer of f on the line through x and y. But y is on the line and $f(y) > f(x)$, a contradiction. The proof of (ii) is the same. //

- How did I know the firm's profit function in our example was concave?

- Theorem: Let $X \subseteq \mathbb{R}^n$ be convex, and let $f: X \to \mathbb{R}$ be twice differentiable with continuous cross partial derivatives.

  (i) f is concave if and only if $D^2 f(x)$ is negative semi-definite for all x in the interior of X.

  (ii) f is convex if and only if $D^2 f(x)$ is positive semi-definite for all x in the interior of X.

The proof again just looks at the graph of f above different lines...

- In the example, the Hessian of the profit function was negative definite for all interior x (in the example, the Hessian was actually independent of x), as long as $p \le 6$. So the profit function was concave, and every critical point was a global maximum!

- I've skipped some stuff on strict concavity and strict convexity, and we haven't touched on a weakening of concavity called "quasi-concavity." You're encouraged to check Simon + Blume's Chapter 21 and Sundaram's Chapters 7 and 8 for these topics.

- Our last topic will be "constrained optimization," which means, basically, the analysis of maximizers on the boundary of X. References are Simon + Blume's Chapters 18 and 19 and Sundaram's Chapters 5 and 6.

- Let's consider the case where

$$X = \{x \in \mathbb{R}^n \mid g(x) = c\},$$

for some differentiable $g: X \to \mathbb{R}$. It may be that g is linear, in which case $g(x) = p \cdot x$ for some $p \in \mathbb{R}^n$.

- Sometimes the maximization problem will be written

$$\max f(x)$$

$$\text{s.t. } g(x) = c,$$

where "s.t." is "subject ~~that~~ to" — indicating that we're only interested in maximizing the value of f over points satisfying $g(x) = c$.

PSC 403

- A point $x \in \mathbb{R}^n$ is a <u>constrained local maximizer</u> of f subject to $g(x)=c$ if there exists some $\varepsilon > 0$ such that, for all $y \in B_\varepsilon(x)$, $g(y)=c$ implies $f(x) \geq f(y)$.

- Theorem of Lagrange: Let $f: \mathbb{R}^n \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}$ be differentiable at x. If x is a constrained local maximizer of f subject to $g(x)=c$ then there exist $\mu, \lambda \in \mathbb{R}$, not both zero, such that

continuously differentiable

$$\mu \nabla f(x) + \lambda \nabla g(x) = 0.$$

- Note: This is saying that the gradient vectors $\nabla f(x)$ and $\nabla g(x)$ are linearly dependent!

- Corollary: Make the same assumptions as above, but add $\nabla g(x) \neq 0$. Then there exists $\lambda \in \mathbb{R}$ such that

Case 1: $\nabla f(x) = 0$
Case 2: $\nabla f(x) \neq 0$

$$\nabla f(x) - \lambda \nabla g(x) = 0,$$

or,

$$\frac{\partial f}{\partial x_1}(x) - \lambda \frac{\partial g}{\partial x_1}(x) = 0$$

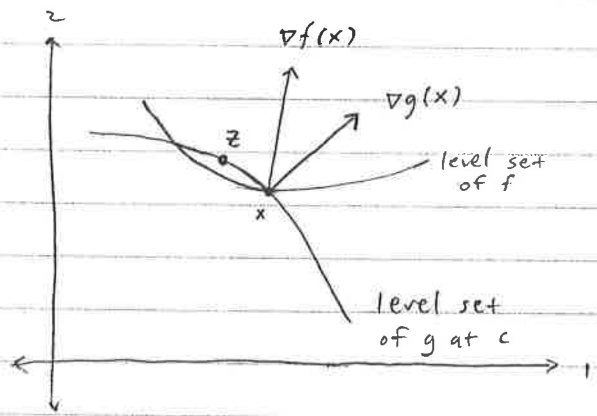
$$\vdots$$

$$\frac{\partial f}{\partial x_n}(x) - \lambda \frac{\partial g}{\partial x_n}(x) = 0$$

- The number, λ, is called the <u>Lagrange Multiplier</u> corresponding to the constraint.

The formal proof uses the implicit function theorem.

- To illustrate the theorem, suppose $\nabla f(x)$ and $\nabla g(x)$ aren't linearly dependent, so the situation looks something like that below, where $n=2$.



Then there will be some point like z such that $g(z)=c$ and z is on a higher level set than x. Moreover, z can be taken arbitrarily close to x, so x can't be a constrained local maximizer.

- Note the implication of the above corollary: for all i and j, as long as $\frac{\partial f}{\partial x_j}(x) \neq 0$ and $\frac{\partial g}{\partial x_j}(x) \neq 0$,

$$\frac{\partial f/\partial x_i(x)}{\partial f/\partial x_j(x)} = \frac{\partial g/\partial x_i(x)}{\partial g/\partial x_j(x)}.$$

When $g(x)=p\cdot x$, this becomes

$$\frac{\partial f/\partial x_i(x)}{\partial f/\partial x_j(x)} = \frac{p_i}{p_j},$$

generalizing the familiar condition, $MRS_{1,2}(x) = \frac{p_1}{p_2}$, from microeconomics.

- There is an easy way to remember the conditions in the corollary to the Theorem of Lagrange: If x is a constrained local maximizer of f subject to $g(x)=c$, and $\nabla g(x) \neq 0$, then there exists $\lambda \in \mathbb{R}$ such that (x,λ) is a critical point of the function $L: \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$L(x,\lambda) = f(x) + \lambda\,(c - g(x)).$$

That is, there is a λ such that

$$\frac{\partial L}{\partial x_1}(x,\lambda) = \frac{\partial f}{\partial x_1}(x) - \lambda \frac{\partial g}{\partial x_1}(x) = 0$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\partial L}{\partial x_n}(x,\lambda) = \frac{\partial f}{\partial x_n}(x) - \lambda \frac{\partial g}{\partial x_n}(x) = 0$$

$$\frac{\partial L}{\partial \lambda}(x,\lambda) = g(x) - c = 0.$$

Clearly, these are the same conditions as in the corollary.

- The function L is called the <u>Lagrangian function</u>.

- Example: Let's go back to the consumer's maximization problem,

$$\max_{x_1, x_2 \geq 0} \; u(x_1, x_2)$$

$$\text{s.t.} \quad p_1 x_1 + p_2 x_2 = I. \qquad g(x_1,x_2) = c \qquad \nabla g(x) = (p_1, p_2)$$

$p_1 > 0 \quad p_2 > 0$

Note: I have written the budget constraint as an equality. This is o.k. as long as the utility function is "monotonic" in this sense: if $y_1 > x_1$ and $y_2 > x_2$ then $u(y_1, y_2) > u(x_1, x_2)$. (Because the consumer will never want to throw money away.) Note that the set $\{(x_1,x_2) \in \mathbb{R}^2 \mid x_1 \geq 0,\ x_2 \geq 0,\ p_1 x_1 + p_2 x_2 = I\}$ is compact. If we assume that u is continuous, the consumer is assured of having a maximizer. If we further assume u is differentiable and the ...

Unlike our example with the firm, this is a very friendly example.

maximizer has $x_1 > 0$ and $x_2 > 0$, then it is a constrained local maximizer of u subject to $p_1 x_1 + p_2 x_2 = I$ and the Theorem of Lagrange applies. The Lagrangian is

$$L(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda (I - p_1 x_1 - p_2 x_2),$$

and the necessary conditions are

$$\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) = \frac{\partial u}{\partial x_1}(x_1, x_2) - \lambda p_1 = 0$$

$$\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = \frac{\partial u}{\partial x_2}(x_1, x_2) - \lambda p_2 = 0$$

$$\frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = I - p_1 x_1 - p_2 x_2 = 0 .$$

Solving these conditions gives us the maximizer — but it may also give us other local (but not global) maximizers and even some local minimizers. If you get more than one solution, you just have to check. If you're lucky, the function u is concave and you don't have to worry.
Let's consider $u(x_1, x_2) = x_1^{\alpha_1} x_2^{\alpha_2}$. Then

$$\alpha_1 x_1^{\alpha_1 - 1} x_2^{\alpha_2} - \lambda p_1 = 0$$

$$\alpha_2 x_1^{\alpha_1} x_2^{\alpha_2 - 1} - \lambda p_2 = 0.$$

Divide to get $\frac{\alpha_1}{\alpha_2} \frac{x_2}{x_1} = \frac{p_1}{p_2}$, or $x_2 = \frac{\alpha_2}{\alpha_1} \frac{p_1}{p_2} x_1$. Plug into $p_1 x_1 + p_2 x_2 = I$ to get

$$p_1 x_1 + p_2 \left(\frac{\alpha_2}{\alpha_1} \frac{p_1}{p_2} x_1 \right) = I$$

or

$$p_1 x_1 + \frac{\alpha_2}{\alpha_1} p_1 x_1 = I$$

PSC 403

or

$$x_1 = \frac{\alpha_1}{\alpha_1+\alpha_2} \frac{I}{p_1}.$$

Going back to our expression for x_2, we get

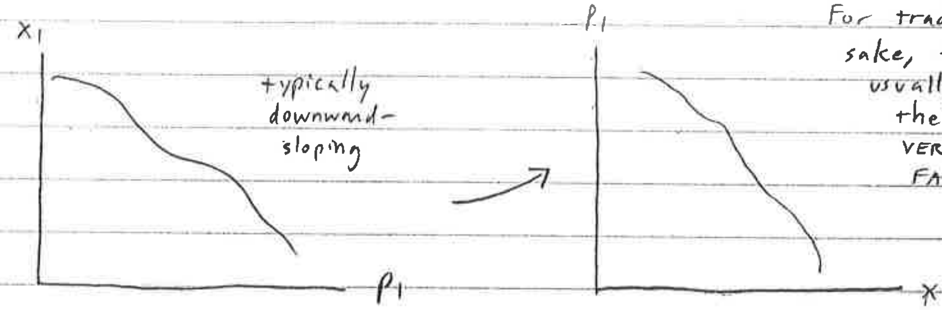
$$x_2 = \frac{\alpha_2}{\alpha_1+\alpha_2} \frac{I}{p_2}$$

The maximizer is, of course, a function of I, p_1, and p_2, so we can write

$$x_1(p_1, p_2, I) = \frac{\alpha_1}{\alpha_1+\alpha_2}\frac{I}{p_1} \quad \text{and} \quad x_2(p_1,p_2,I) = \frac{\alpha_2}{\alpha_1+\alpha_2}\frac{I}{p_2}.$$

Note: this is most straightforward when, as in this example, the consumer's maximizer is unique.

In consumer theory, these are called the consumer's <u>demand functions</u>. (Note: the dependence on α_1 and α_2 is typically suppressed.) Fixing p_2 and I, we can graph x_1 as a function of one variable, p_1. This gives us the <u>demand curve</u> for good 1.



For tradition's sake, economists usually switch the axes — A VERY FRUSTRATING FACT.

(Similarly for good 2.) Taking partial derivatives,

$$\frac{\partial x_1}{\partial p_1}, \quad \frac{\partial x_1}{\partial p_2}, \quad \frac{\partial x_1}{\partial I}, \quad \frac{\partial x_2}{\partial p_1}, \quad \frac{\partial x_2}{\partial p_2}, \quad \frac{\partial x_2}{\partial I},$$

tells us how the consumer's maximizer changes with respect to market parameters. This is called <u>comparative statics</u>.

PSC 403

– The last thing I want to cover (I'll never get this far in class, though) is the interpretation of the Lagrange multiplier. Let's assume, for simplicity, there is a unique maximizer. Write $x_1(c), x_2(c), \ldots, x_n(c), \lambda(c)$ for the solution to the Lagrangian problem given c, and define

$$h(c) = f(x_1(c), x_2(c), \ldots, x_n(c))$$

to be the highest attainable value of f subject to $g(x_1, \ldots x_n) = c$. Assuming that $x_1(\cdot), \ldots, x_n(\cdot)$ are differentiable, I claim

$$\frac{dh}{dc}(c) = \lambda(c).$$

Note: when $\lambda = 0$, i.e., $\nabla f(x) = 0$, the constraint is "non-binding"; it is "binding" otherwise.

That is, λ tells us the rate at which the maximized value of f would change if c were increased. In the consumer's problem, it tells us how the consumer's utility would be increased if he or she had more money. The proof is quite easy. Note that, for all c,

$$① \quad g(x_1(c), x_2(c), \ldots, x_n(c)) = c,$$

and

$$② \quad \frac{\partial f}{\partial x_i}(x_1(c), \ldots, x_n(c)) - \lambda(c) \frac{\partial g}{\partial x_i}(x_1(c), \ldots, x_n(c)) = 0$$

for all $i = 1 \ldots n$. Differentiating ① and using the chain rule,

$$③ \quad \frac{\partial g}{\partial x_1}\frac{dx_1}{dc}(c) + \frac{\partial g}{\partial x_2}\frac{dx_2}{dc}(c) + \ldots + \frac{\partial g}{\partial x_n}\frac{dx_n}{dc}(c) = 1,$$

where each $\frac{\partial g}{\partial x_i}$ is evaluated at $x_1(c), x_2(c), \ldots x_n(c)$.

PSC 403

Now differentiating h and using the chain rule,

$$\frac{dh}{dc}(c) = \frac{\partial f}{\partial x_1}\frac{dx_1}{dc}(c) + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dc}(c).$$

By ②, we can subtract $\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i}$ to the righthand side, for $i = 1, \ldots, n$.

$$\begin{aligned}\frac{dh}{dc}(c) &= \lambda(c)\frac{\partial g}{\partial x_1} + \cdots + \lambda(c)\frac{\partial g}{\partial x_n} \\ &= \lambda(c)\left(\frac{\partial g}{\partial x_1} + \cdots + \frac{\partial g}{\partial x_n}\right) \\ &= \lambda(c),\end{aligned}$$

where the last step uses ③.

– Some things we didn't get to:

- maximization problems with multiple equality constraints
- maximization problems with inequality constraints
- problems with both equality and inequality constraints
- second order conditions for constrained problems, bordered Hessians
- quasi-concave functions and maximization problems
- iterated integrals

– Those things aside, you guys learned a lot of math. Good work, and have a good break.

J.D.