Lecture 3
Linear algebra review

- vector space, subspaces
- independence, basis, dimension
- range, nullspace, rank
- change of coordinates
- norm, angle, inner product

Vector spaces

a *vector space* or *linear space* (over the reals) consists of

- a set $\mathcal{V}$
- a vector sum $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$
- a scalar multiplication : $\mathbf{R} \times \mathcal{V} \to \mathcal{V}$
- a distinguished element $0 \in \mathcal{V}$

which satisfy a list of properties

- $x + y = y + x, \quad \forall x, y \in \mathcal{V}$ $\quad$ (+ is commutative)
- $(x + y) + z = x + (y + z), \quad \forall x, y, z \in \mathcal{V}$ $\quad$ (+ is associative)
- $0 + x = x$, $\forall x \in \mathcal{V}$ $\quad$ (0 is additive identity)
- $\forall x \in \mathcal{V} \;\; \exists (-x) \in \mathcal{V}$ s.t. $x + (-x) = 0$ $\quad$ (existence of additive inverse)
- $(\alpha\beta)x = \alpha(\beta x), \quad \forall \alpha, \beta \in \mathbf{R} \quad \forall x \in \mathcal{V}$ $\quad$ (scalar mult. is associative)
- $\alpha(x + y) = \alpha x + \alpha y, \quad \forall \alpha \in \mathbf{R} \;\; \forall x, y \in \mathcal{V}$ $\quad$ (right distributive rule)

this is equality in Right

- $(\alpha + \beta)x = \alpha x + \beta x, \quad \forall \alpha, \beta \in \mathbf{R} \quad \forall x \in \mathcal{V}$ $\quad$ (left distributive rule)

this is a plus in R this plus is a plus of vectors

- $1x = x, \quad \forall x \in \mathcal{V}$

Examples

- $\mathcal{V}_1 = \mathbf{R}^n$, with standard (componentwise) vector addition and scalar multiplication

- $\mathcal{V}_2 = \{0\}$ (where $0 \in \mathbf{R}^n$)

- $\mathcal{V}_3 = \text{span}(v_1, v_2, \ldots, v_k)$ where

$$\text{span}(v_1, v_2, \ldots, v_k) = \{\alpha_1 v_1 + \cdots + \alpha_k v_k \mid \alpha_i \in \mathbf{R}\}$$

span is all possible linear combinations

and $v_1, \ldots, v_k \in \mathbf{R}^n$

Subspaces

- a *subspace* of a vector space is a *subset* of a vector space which is itself a vector space
- roughly speaking, a subspace is closed under vector addition and scalar multiplication

  —subspace must go through the origin so that the scalar multiplication by 0 exists
- examples $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$ above are subspaces of $\mathbf{R}^n$

Vector spaces of functions

- $\mathcal{V}_4 = \{x : \mathbf{R}_+ \to \mathbf{R}^n \mid x \text{ is differentiable}\}$, where vector sum is sum of functions:

  plus in vectors of V4 plus in vectors in Rn

  $$(x+z)(t) = x(t) + z(t)$$

  this is a vector in V4 (x+z)

  and scalar multiplication is defined by

  $$(\alpha x)(t) = \alpha x(t)$$

  (a *point* in $\mathcal{V}_4$ is a *trajectory* in $\mathbf{R}^n$)

- $\mathcal{V}_5 = \{x \in \mathcal{V}_4 \mid \dot{x} = Ax\}$
  (*points* in $\mathcal{V}_5$ are *trajectories* of the linear system $\dot{x} = Ax$)

- $\mathcal{V}_5$ is a subspace of $\mathcal{V}_4$

  — not subset but supspace

Independent set of vectors

a set of vectors $\{v_1, v_2, \ldots, v_k\}$ is *independent* if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = 0$$

independence is an attribute of a set of vectors, not of vectors

some equivalent conditions:

- coefficients of $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ are uniquely determined, *i.e.*,

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_k v_k$$

  implies $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_k = \beta_k$

- no vector v_i can be expressed as a linear combination of the other vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k$

  — this comes directly from the definition of independance

Basis and dimension

basis is an attribute of a set of vectors

set of vectors $\{v_1, v_2, \ldots, v_k\}$ is a *basis* for a vector space $\mathcal{V}$ if

- $v_1, v_2, \ldots, v_k$ span $\mathcal{V}$, *i.e.*, $\mathcal{V} = \mathrm{span}(v_1, v_2, \ldots, v_k)$
- $\{v_1, v_2, \ldots, v_k\}$ is independent

equivalent: every $v \in \mathcal{V}$ can be uniquely expressed as

$$v = \alpha_1 v_1 + \cdots + \alpha_k v_k$$

fact: for a given vector space $\mathcal{V}$, the number of vectors in any basis is the same

a vector space can have multiple basis, but they each have to have the same number of elements (same dimension)

number of vectors in any basis is called the *dimension* of $\mathcal{V}$, denoted $\mathbf{dim}\mathcal{V}$

(we assign $\mathbf{dim}\{0\} = 0$, and $\mathbf{dim}\mathcal{V} = \infty$ if there is no basis)

Nullspace of a matrix

the *nullspace* of $A \in \mathbf{R}^{m \times n}$ is defined as

$$\mathcal{N}(A) = \{ \, x \in \mathbf{R}^n \mid Ax = 0 \, \}$$ this is a subspace of R^n

— columns of A being linearly independent, makes the nullspace of A = 0

- $\mathcal{N}(A)$ is set of vectors mapped to zero by $y = Ax$
- $\mathcal{N}(A)$ is set of vectors orthogonal to all rows of A

$\mathcal{N}(A)$ gives *ambiguity* in x given $y = Ax$:

- if $y = Ax$ and $z \in \mathcal{N}(A)$, then $y = A(x + z)$ because A*z = 0
- conversely, if $y = Ax$ and $y = A\tilde{x}$, then $\tilde{x} = x + z$ for some $z \in \mathcal{N}(A)$

Zero nullspace

A is called *one-to-one* if 0 is the only element of its nullspace:
$\mathcal{N}(A) = \{0\} \iff$

- x can always be uniquely determined from $y = Ax$
  ($i.e.$, the linear transformation $y = Ax$ doesn't 'lose' information)

- mapping from x to Ax is one-to-one: different x's map to different y's

- columns of A are independent (hence, a basis for their span)

Ax=0
— the only way to make this work is to have each of the coefficients = 0
.: A is independent

- A has a *left inverse*, $i.e.$, there is a matrix $B \in \mathbf{R}^{n \times m}$ s.t. $BA = I$

I is dimension nx

- $\det(A^T A) \neq 0$

for the last two bullet points look at the lecture video for explanation

(we'll establish these later)

Interpretations of nullspace

listen to the video for this slide too

suppose $z \in \mathcal{N}(A)$

$y = Ax$ represents **measurement** of x

- z is undetectable from sensors — get zero sensor readings
- x and $x + z$ are indistinguishable from sensors: $Ax = A(x + z)$

$\mathcal{N}(A)$ characterizes *ambiguity* in x from measurement $y = Ax$

$y = Ax$ represents **output** resulting from input x

- z is an input with no result
- x and $x + z$ have same result

$\mathcal{N}(A)$ characterizes *freedom of input choice* for given result

Range of a matrix

the *range* of $A \in \mathbf{R}^{m \times n}$ is defined as

$$\mathcal{R}(A) = \{Ax \mid x \in \mathbf{R}^n\} \subseteq \mathbf{R}^m$$ (output space)

$\mathcal{R}(A)$ can be interpreted as

- the set of vectors that can be 'hit' by linear mapping $y = Ax$
- the span of columns of A
- the set of vectors y for which $Ax = y$ has a solution

  the set of vectors which can be generated using A

Onto matrices

A is called *onto* if $\mathcal{R}(A) = \mathbf{R}^m \iff$

- $Ax = y$ can be solved in x for any y
- columns of A span $\mathbf{R}^m$
- A has a *right inverse*, $i.e.$, there is a matrix $B \in \mathbf{R}^{n \times m}$ s.t. $AB = I$
- rows of A are independent
- $\mathcal{N}(A^T) = \{0\}$
- $\det(AA^T) \neq 0$

(some of these are not obvious; we'll establish them later)

Interpretations of range

suppose $v \in \mathcal{R}(A)$, $w \notin \mathcal{R}(A)$

$y = Ax$ represents **measurement** of x

- $y = v$ is a *possible* or *consistent* sensor signal Y=Ax works and we can generate v
- $y = w$ is *impossible* or *inconsistent*; sensors have failed or model is wrong Y=Ax doesnt work anymore and we cannot generate w

$y = Ax$ represents **output** resulting from input x

- v is a possible result or output
- w cannot be a result or output

listen to the example on this slide

$\mathcal{R}(A)$ characterizes the *possible results* or *achievable outputs*

Inverse

$A \in \mathbf{R}^{n \times n}$ is *invertible* or *nonsingular* if $\det A \neq 0$

square matrix

equivalent conditions:

- columns of A are a basis for $\mathbf{R}^n$
- rows of A are a basis for $\mathbf{R}^n$
- $y = Ax$ has a unique solution x for every $y \in \mathbf{R}^n$
- A has a (left and right) inverse denoted $A^{-1} \in \mathbf{R}^{n \times n}$, with $AA^{-1} = A^{-1}A = I$

  BA=I
  AB=I
- $\mathcal{N}(A) = \{0\}$
- $\mathcal{R}(A) = \mathbf{R}^n$
- $\det A^T A = \det AA^T \neq 0$

  det(A^T)*det(A) =/= 0
  — because det(A) is not 0 (shown above)

Interpretations of inverse

suppose $A \in \mathbf{R}^{n \times n}$ has inverse $B = A^{-1}$

- mapping associated with B undoes mapping associated with A (applied either before or after!)
- $x = By$ is a perfect (pre- or post-) *equalizer* for the *channel* $y = Ax$
- $x = By$ is unique solution of $Ax = y$

Dual basis interpretation

watch this slide

- let a_i be columns of A, and $\tilde{b}_i^T$ be rows of $B = A^{-1}$

- from $y = x_1 a_1 + \cdots + x_n a_n$ and $x_i = \tilde{b}_i^T y$, we get

bi^T*aj = 0 i=/= j
= 1 i=j

$$y = \sum_{i=1}^{n} (\tilde{b}_i^T y) a_i$$

thus, inner product with *rows of inverse matrix* gives the coefficients in the *expansion of a vector in the columns of the matrix*

- $\{\tilde{b}_1, \ldots, \tilde{b}_n\}$ and $\{a_1, \ldots, a_n\}$ are called *dual bases*

Rank of a matrix

we define the *rank* of $A \in \mathbf{R}^{m \times n}$ as

$$\mathbf{rank}(A) = \mathbf{dim}\, \mathcal{R}(A)$$

(nontrivial) **facts:**

- $\mathbf{rank}(A) = \mathbf{rank}(A^T)$
- $\mathbf{rank}(A)$ is maximum number of independent columns (or rows) of A hence $\mathbf{rank}(A) \leq \mathbf{min}(m, n)$ rank cant be bigger than height or width of matrix
- $\mathbf{rank}(A) + \mathbf{dim}\, \mathcal{N}(A) = n$

Conservation of dimension

interpretation of **rank**$(A) +$ **dim** $\mathcal{N}(A) = n$:

- **rank**(A) is dimension of set 'hit' by the mapping $y = Ax$
- **dim** $\mathcal{N}(A)$ is dimension of set of x 'crushed' to zero by $y = Ax$
- 'conservation of dimension': each dimension of input is either crushed to zero or ends up in output

Ex:
If you have ten knobs affecting an output. If you have rank 8, then your output will be only be 8 dimensional (as opposed to the expected 10).
This
means that the dimension of the null space is 2. So all the knobs work, but if you were to stop using the 2 knobs in the null space, you would be

- roughly speaking:
  - n is number of degrees of freedom in input x
  - **dim** $\mathcal{N}(A)$ is number of degrees of freedom lost in the mapping from x to $y = Ax$
  - **rank**(A) is number of degrees of freedom in output y

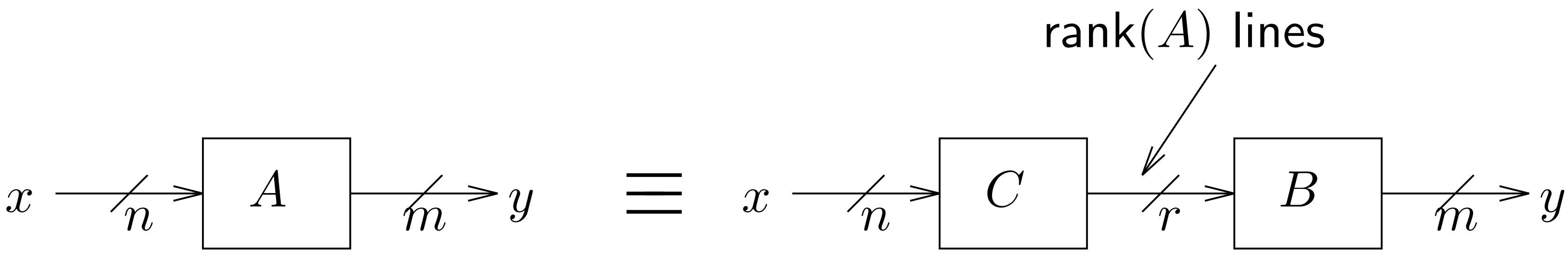
‘Coding’ interpretation of rank

- rank of product: $\textbf{rank}(BC) \le \textbf{min}\{\textbf{rank}(B),\ \textbf{rank}(C)\}$

- hence if $A = BC$ with $B \in \mathbf{R}^{m\times r}$, $C \in \mathbf{R}^{r\times n}$, then $\textbf{rank}(A) \le r$

  as well as rank(A)<= m
  and rank(A) <= n

- conversely: if $\textbf{rank}(A) = r$ then $A \in \mathbf{R}^{m\times n}$ can be factored as $A = BC$ with $B \in \mathbf{R}^{m\times r}$, $C \in \mathbf{R}^{r\times n}$:



- $\textbf{rank}(A) = r$ is minimum size of vector needed to faithfully reconstruct y from x

Application: fast matrix-vector multiplication

view this slide in video

- need to compute matrix-vector product $y = Ax$, $A \in \mathbf{R}^{m \times n}$
- A has known factorization $A = BC$, $B \in \mathbf{R}^{m \times r}$
- computing $y = Ax$ directly: mn operations
- computing $y = Ax$ as $y = B(Cx)$ (compute $z = Cx$ first, then $y = Bz$): $rn + mr = (m+n)r$ operations
- savings can be considerable if $r \ll \min\{m, n\}$

Full rank matrices

watch this slide

for $A \in \mathbf{R}^{m \times n}$ we always have $\mathbf{rank}(A) \leq \mathbf{min}(m, n)$

we say A is *full rank* if $\mathbf{rank}(A) = \mathbf{min}(m, n)$

- for **square** matrices, full rank means nonsingular
- for **skinny** matrices ($m \geq n$), full rank means columns are independent
- for **fat** matrices ($m \leq n$), full rank means rows are independent

Change of coordinates

watch all of these slides

'standard' basis vectors in $\mathbf{R}^n$: $(e_1, e_2, \ldots, e_n)$ where

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

(1 in ith component)

obviously we have

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$$

x_i are called the coordinates of x (in the standard basis) if $(t_1, t_2, \ldots, t_n)$ is another basis for $\mathbf{R}^n$, we have

$$x = \tilde{x}_1 t_1 + \tilde{x}_2 t_2 + \cdots + \tilde{x}_n t_n \quad {\scriptstyle =T\tilde{x}}$$

where $\tilde{x}_i$ are the coordinates of x in the basis $(t_1, t_2, \ldots, t_n)$

define $T = \begin{bmatrix} t_1 & t_2 & \cdots & t_n \end{bmatrix}$ so $x = T\tilde{x}$, hence

$$\tilde{x} = T^{-1}x$$

(T is invertible since t_i are a basis)

T^{-1} transforms (standard basis) coordinates of x into t_i-coordinates

inner product ith row of T^{-1} with x extracts t_i-coordinate of x

consider linear transformation $y = Ax$, $A \in \mathbf{R}^{n \times n}$

express y and x in terms of $t_1, t_2 \ldots, t_n$:

$$x = T\tilde{x}, \quad y = T\tilde{y}$$

so

$$\tilde{y} = (T^{-1}AT)\tilde{x}$$

- $A \longrightarrow T^{-1}AT$ is called *similarity transformation*
- similarity transformation by T expresses linear transformation $y = Ax$ in coordinates $t_1, t_2, \ldots, t_n$

(Euclidean) norm

for $x \in \mathbf{R}^n$ we define the (Euclidean) norm as

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

$\|x\|$ measures length of vector (from origin)

important properties:

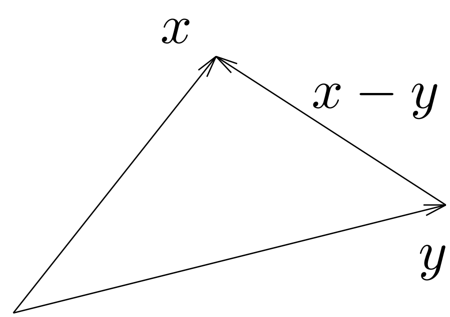
- $\|\alpha x\| = |\alpha|\|x\|$ (homogeneity)
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ (nonnegativity)
- $\|x\| = 0 \iff x = 0$ (definiteness)

RMS value and (Euclidean) distance

root-mean-square (RMS) value of vector $x \in \mathbf{R}^n$:

$$\mathbf{rms}(x) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 \right)^{1/2} = \frac{\|x\|}{\sqrt{n}}$$

norm defines distance between vectors: $\mathbf{dist}(x, y) = \|x - y\|$



Inner product

$$\langle x, y\rangle := x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^Ty$$

important properties:

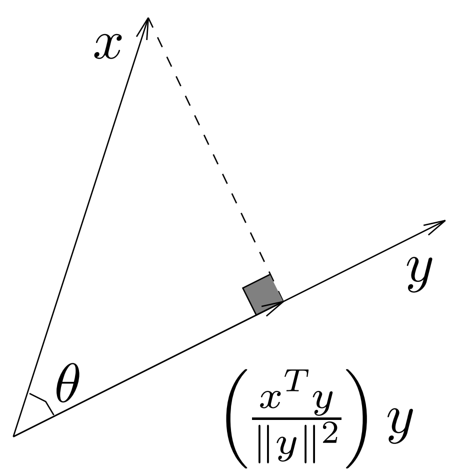
- $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$
- $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
- $\langle x, y\rangle = \langle y, x\rangle$
- $\langle x, x\rangle \geq 0$
- $\langle x, x\rangle = 0 \Longleftrightarrow x = 0$

$f(y) = \langle x, y\rangle$ is linear function : $\mathbf{R}^n \to \mathbf{R}$, with linear map defined by row vector x^T

Cauchy-Schwarz inequality and angle between vectors

- for any $x, y \in \mathbf{R}^n$, $|x^T y| \leq \|x\|\|y\|$
- (unsigned) angle between vectors in $\mathbf{R}^n$ defined as

$$\theta = \angle(x, y) = \cos^{-1} \frac{x^T y}{\|x\|\|y\|}$$



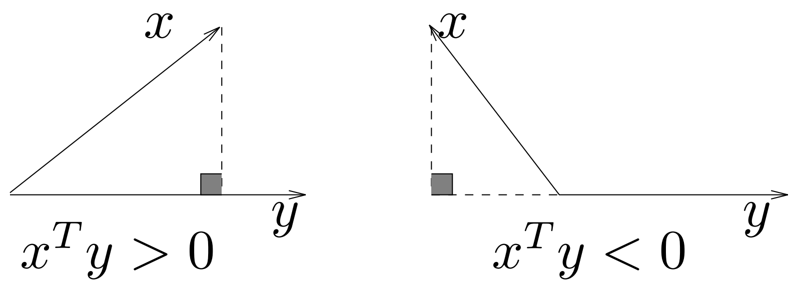
thus $x^T y = \|x\|\|y\| \cos\theta$

special cases:

- x and y are *aligned*: $\theta = 0$; $x^T y = \|x\|\|y\|$;
  (if $x \neq 0$) $y = \alpha x$ for some $\alpha \geq 0$
- x and y are *opposed*: $\theta = \pi$; $x^T y = -\|x\|\|y\|$
  (if $x \neq 0$) $y = -\alpha x$ for some $\alpha \geq 0$
- x and y are *orthogonal*: $\theta = \pi/2$ or $-\pi/2$; $x^T y = 0$
  denoted $x \perp y$

interpretation of $x^T y > 0$ and $x^T y < 0$:

- $x^T y > 0$ means $\angle(x, y)$ is acute
- $x^T y < 0$ means $\angle(x, y)$ is obtuse



$\{x \mid x^T y \leq 0\}$ defines a *halfspace* with outward normal vector y, and boundary passing through 0