LINEAR ALGEBRA Cliff Notes

- We'll work in $\mathbb{R}^n$ usually (or an isomorphic space like $\mathbb{R}^{m\times n}$) and use its <u>vector space</u> properties.

- Measure "size" w/ a <u>norm</u> $\|\cdot\|$. Let V be a vector space

  <u>Def</u> a norm $\|\cdot\| : V \to \mathbb{R}_+$ ($\mathbb{R}_+$ = non-negative $\mathbb{R}$ numbers)

  satisfies
  1. $\|x\| = 0$ iff $x = 0$
  2. $\forall \alpha \in \mathbb{F}$, $\|\alpha x\| = |\alpha| \cdot \|x\|$, ($\mathbb{F}$ is a field, for us $\mathbb{R}$ or $\mathbb{C}$)
  3. $\|x+y\| \le \|x\| + \|y\|$ "triangle ineq." or "Δ-ineq."

- Measure "distance" w/ metric induced by norm

  $d(x,y) = \|x-y\|$.

  <u>Fact</u>: reverse Δ-ineq: $\|x\| = \|x+y-y\| \le \|x+y\| + \|y\|$

  $\|x\| = \|x-y+y\| \le \|x-y\| + \|y\|$

  so ① $\|x+y\| \ge \|x\| - \|y\|$

  and ② $\|x+y\| \ge \|y\| - \|x\|$ by interchanging x,y

  ③ $\|x-y\| \ge \|x\| - \|y\|$ ④ $\|x-y\| \ge \|y\| - \|x\|$

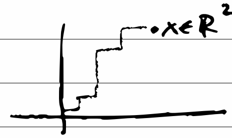
<u>Ex</u>: Euclidean norm/metric, $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ for $x \in \mathbb{R}^n$

"distance as the crow flies"

<u>Ex</u>: Mahalanobis $\|x\|_V = x^T V x$ for a matrix $\underbrace{V > 0}_{\text{pos. def.}}$

<u>Ex</u>: Taxi-cab/Manhattan distance (i.e. ℓ_1 norm)

$\|x\|_1 = \sum |x_i|$



<u>Inner product</u>: For a vector space $(V, \mathbb{F})$, $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$

"angles" is bi-linear (well, almost, if have $\mathbb{C}$-numbers)

Usually use Euclidean inner product $\langle x, y\rangle = x^T y = \sum_i x_i y_i$

which induces $\|\cdot\|_2$ ($\|x\|_2^2 = \langle x, x\rangle$)

<u>Fact</u>: $\|\cdot\|_p$ ($p \ne 2$) norms not induced by any inner product

if $V = \mathbb{C}^n$, two options: (1) $\mathbb{F} = \mathbb{R}$, then output of $\langle \cdot, \cdot \rangle$ must be $\mathbb{R}$

So $\langle x, y \rangle = \mathrm{Re}(x^* y)$

x^* means $\bar{x}^T$ ← transpose (complex conjugate)

adjoint

(2) $\mathbb{F} = \mathbb{C}$, then $\langle x, y \rangle = x^* y$

Positive definite matrix

$A > 0$ means A is p.d., meaning $\forall x \neq 0$, $\langle x, Ax \rangle > 0$

Usually we implicitly impose that $A = A^*$

in which case $A \succ 0$ iff all eigenvalues $\lambda_i > 0$

"squiggly >"

and pos. semidefinite (psd) iff all $\lambda_i \geq 0$.

Recall a matrix is diagonalizable if $A = VDV^{-1}$
V invertible
D diag.

$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ not diagonalizable.

If A is normal ($AA^* = A^*A$) $\Rightarrow$ it is unitarily diagonalizable

$A = VDV^{-1} = VDV^*$, $V^{-1} = V^*$

If A is self-adjoint / Hermitian ($A = A^*$) $\Rightarrow$ it is normal and $D_{ii} = \lambda_i \in \mathbb{R}$

Matrix norms

- Frobenius / Hilbert-Schmidt $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \langle A, A \rangle$

where $\langle A, A \rangle = \mathrm{tr}(A^* A) = \mathrm{vec}(A)^* \mathrm{vec}(A)$

$\left.\begin{array}{l} \mathrm{vec}: \mathbb{R}^{m \times n} \to \mathbb{R}^{mn \times 1} \\ \mathrm{mat}: \mathbb{R}^{mn \times 1} \to \mathbb{R}^{m \times n} \end{array}\right\}$ inverses and adjoints of each other

- Operator norm

$$\|A\|_{a \to b} := \sup_{x \neq 0} \frac{\|Ax\|_b}{\|x\|_a} = \sup_{\|x\|_a = 1} \|Ax\|_b$$

Submultiplicative: $\|Ax\|_b \leq \|A\|_{a \to b} \cdot \|x\|_a$

the most common operator norm is the "spectral norm" using $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$ (Eucl. norm)

$\|A\|_{2\to 2}$, usually written $\|A\|_2$ or just $\|A\|$

⚠ in Matlab, norm(A) is spectral norm by default
in Python, norm(A) is Frobenius norm by default

$\|A\|_2$ calculated via SVD: $A = U \Sigma V^*$
$\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$ then $\|A\|_2 = \|\sigma\|_\infty$

$\|A\|_F = \|\sigma\|_2$

If A is $n \times n$, $\|A\|_2$ requires $O(n^3)$ flops
$\|A\|_F$ requires $O(n^2)$ flops

In finite dimensions (like $\mathbb{R}^n$), all norms are equivalent, which means for norms $\|\cdot\|_a$, $\|\cdot\|_b$, $\exists\, c, C > 0$ s.t.

$$\forall x \in \mathbb{R}^n, \quad c\|x\|_a \le \|x\|_b \le C \cdot \|x\|_a$$

Equivalent $\neq$ the same. Only means they induce the same topology

Open sets: $\mathcal{U} \subseteq \mathbb{R}^n$ is open if $\forall\, x \in \mathcal{U}$, $\exists\, \varepsilon > 0$ s.t. the open ball $B_\varepsilon(x) \subseteq \mathcal{U}$

Closed set: $F \subseteq \mathbb{R}^n$ closed if F^c is open (complement)

or, equiv., if it is sequentially closed, $(x_n) \subseteq F$
$x_n \to x \Rightarrow x \in F$. "Contains its limit points"

Compact: since in $\mathbb{R}^n$, means closed and bounded
or, equiv., if $(x_n) \subseteq K$, K compact, then $\exists$ a convergent subseq. of (x_n) w/ limit in K

More misc. notes:

A symmetric matrix X is positive semi-definite ($X \succeq 0$)
iff $\exists$ matrix G s.t. $X = GG^T$ (or GG^* if $\mathbb{C}$)

(if $X \succ 0$, then G is square & invertible and called the Cholesky factor of X)

"$\succeq$" induces a partial order on S^n, called the "Loewner order"
not a total order:
not all elements comparable
(but if they are, then acts nicely, e.g.
$A \succeq B,\ B \succeq C \Rightarrow A \succeq C$ etc.)

to say $A \succeq B$ means $A - B \succeq 0$

⚠ If A has eigenvalues λ_i
and B has eigenvalues σ_i
then $A-B$ doesn't necessarily have eigenvalues $\lambda_i - \sigma_i$

trace

$$tr(A) = \sum_{i=1}^{n} A_{ii} \quad \text{is sum of diagonal entries}$$
$$= \sum_i \lambda_i \quad \text{sum of eigenvalues}$$

$tr(ABC) = tr(BCA) = tr(CAB)$ "cyclic property"

tricks:

- Compute $tr(A^T B)$, $A, B \in \mathbb{R}^{n \times n}$

BAD: { (1) compute $C = A^T B$ — $O(n^3)$ flops
(2) find $tr(C)$ — $O(n)$ flops }

BETTER: $tr(A^T B) = \langle A, B \rangle = \langle vec(A), vec(B) \rangle$
$= vec(A)^T vec(B)$
$O(n^2)$ flops

- Compute $\|A\|_F$, A is $n \times n$

BAD: $\|A\|_F = \sqrt{tr(A^T A)}$ so compute $A^T A$ ... $O(n^3)$

BAD: Find SVD of A, $\|A\|_F^2 = \sum_i \sigma_i^2$ — $O(n^3)$

GOOD: $= \sqrt{\sum_i \sum_j A_{ij}^2}$ — $O(n^2)$

- Fancier trick. Let A be low-rank, $A = uv^T$

$$\underset{m\times n}{A} = \underset{m\times k}{u}\ \underset{k\times n}{v^T}$$

For simplicity, let it be rank 1 and square: $k=1$, $m=n$

Suppose B is sparse, $n \times n$

Compute $\|A-B\|_F$ efficiently.

Naive: $\sqrt{\sum_{i,j} (A_{ij}-B_{ij})^2}$ $O(n^2)$ ← flops, and also $O(n^2)$ memory

Better:

$$\|A-B\|_F^2 = \underset{①}{\|A\|_F^2} + \underset{②}{\|B\|_F^2} - 2\,\underset{③}{tr(A^T B)}$$

① $\|A\|_F^2 = \|uv^T\|^2 = tr((uv^T)^T uv^T)$

$= tr(vu^Tuv^T)$

$= tr(\underbrace{u^Tu}_{\in\mathbb{R}}\, v^Tv)$) cyclic $= u^Tu \cdot v^Tv$

$\in\mathbb{R}$, costs $O(n)$ to compute ✓

② $\|B\|_F^2$ costs $nnz(B)$ to compute (hopefully $\ll O(n^2)$)

③ $tr(A^TB) = tr((uv^T)^TB) = tr(vu^TB)$

$= tr(u^TBv)$

compute $w = Bv$ (sparse matrix-vector multiply) in $nnz(B)$ time, $tr(u^Tw) = u^Tw$ $O(n)$ time

so $O(n + nnz(B))$ time and memory.

- Matrix inversion lemma (ie. Sherman-Morrison-Woodbury)

Easy (=quick) to invert a low-rank perturbation of a matrix for which you already know the inverse

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}$$

and see wikipedia for generalizations

Basically exploiting the Schur Complement

· Hölder's inequality

$$|\langle x,y\rangle| \leq \|x\|_p \cdot \|y\|_q \quad \text{if} \quad \underbrace{\frac{1}{p}+\frac{1}{q}}^{\text{Hölder conjugates}} = 1$$

e.g., $p=q=2$ "Cauchy-Schwarz"

e.g., $p=1$, $q=\infty$ or vice-versa

· Norms, inner products are continuous

① $\lim \|x_n\| = \|\lim x_n\|$ if $\lim x_n$ exists

② $\lim \langle a, x_n\rangle = \langle a, \lim x_n\rangle$ if $\lim x_n$ exists

· If $x_n \to x$ and $y_n \to y$, then

$\lim_n \langle x_n, y_n\rangle = \langle x, y\rangle$

proof: $y_n \to y \Rightarrow (y_n)$ bounded, i.e. $\|y_n\| \leq M \ \forall n$

$$\lim_n \langle x_n, y_n\rangle = \underbrace{\lim_n \langle x_n - x, y_n\rangle} + \underbrace{\lim_n \langle x, y_n\rangle}_{=\langle x,y\rangle \text{ via } ②}$$

$|\lim \langle x_n - x, y_n\rangle|$

$= \lim |\langle x_n - x, y_n\rangle|$ via ①

$\leq \lim \|x_n - x\| \cdot \|y_n\|$

$\leq M \cdot \lim \|x_n - x\| = 0$