Review of Linear Algebra

Vector and matrix norms
Basic matrix decompositions
Condition numbers

Maria Cameron, August 31

Vector Spaces

Definition 1. *A vector space V is a set closed with respect to the operations of addition "+": $V \times V \to V$, and multiplication by a scalar "$\cdot\alpha$": $V \to V$ The operations satisfy the following properties.*

(1) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$,

(2) $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$,

(3) $\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$,

(4) $(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$,

(5) there is $0 \in V$ s.t. $\mathbf{a} + \mathbf{0} = \mathbf{a}$ for any $\mathbf{a} \in V$,

(6) for any $\mathbf{a} \in V$ there is $(-\mathbf{a}) \in V$ s.t. $\mathbf{a} + (-\mathbf{a}) = 0$,

(7) $\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$,

(8) $1\mathbf{a} = \mathbf{a}$ for any $\mathbf{a} \in V$.

Exercise Prove that for any $\mathbf{a} \in V$ $0\mathbf{a} = \mathbf{0}$ where $0 \in \mathbb{R}$ while $\mathbf{0} \in V$.

Reminder of basic concepts

- A ***subspace*** ***W*** of a vector space ***V*** = a subset of V closed under addition and scalar multiplication
- The ***span*** of $v_1, \ldots, v_n$ = set of all their linear combinations
- $v_1, \ldots, v_n$ are ***linearly independent*** if any their zero linear combination has all coefficients zero.
- A ***basis*** of ***V*** is a subset of vectors $\{v_j\}_{j \in \mathcal{I}}$ such that
  - any ***v*** is represented, i.e., $v = \sum_{j \in \mathcal{I}} \alpha_j v_j$
  - the set $\{v_j\}_{j \in \mathcal{I}}$ minimal, i.e.,

$$\forall m \in \mathcal{I} \;\; \exists v \in V : \quad v - \sum_{j \in \mathcal{I} \setminus \{m\}} \alpha_j v_j \neq 0$$

- A ***linear transformation*** of a vector space ***V*** to a vector space ***W*** is a map ***L: V —> W*** such that

$$\forall v_1, v_2 \in V \ \forall \alpha_1, \alpha_2 \in \mathbb{R}: \quad L(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 L(v_1) + \alpha_2 L(v_2)$$

- Let $\mathcal{B} = \{b_j\}$ and $\mathcal{E} = \{e_j\}$ be bases in ***V*** and ***W***, respectively.

$$L(v) = L\left(\sum_j v_j b_j\right) = \sum_j v_j L(b_j) = \sum_j v_j \sum_i a_{ij} e_j$$

$A = (a_{ij}) = {}_{\mathcal{E}}[L]_{\mathcal{B}}$ = ***matrix of L*** w.r.t. these bases.

- ***Matrix product:*** $\underbrace{A}_{m\times n}\underbrace{B}_{n\times p} = \underbrace{C}_{m\times p}$

- Matrix ***transpose***: $A^\top = (a_{ij})$; ***adjoint***: $A^* = (\bar{a}_{ij})$

Examples: finite-dimensional vector spaces

- $\mathbb{R}^n$ With the standard basis $\{e_j\}_{j=1}^n$.
- $W := \{v \in V \mid \Sigma_j v_j = 0\}$ is a subspace;

  $W_1 := \{v \in V \mid \Sigma_j v_j = 1\}$ is NOT a subspace
- $\mathcal{P}_n = \{$ polynomials of degree $\leq n\}$ = (n+1)-dimensional vector space. $\mathcal{X} := \{1, x, \ldots, x^n\}$ is a basis.
- $\dfrac{d}{dx} : \mathcal{P}_n \to \mathcal{P}_{n-1}$ is a linear transformation.

$$\underset{n \times (n+1)}{D_{\mathcal{X}}} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 2 & \ldots & \\ & & & \ddots & \\ 0 & & & 0 & n \end{bmatrix}$$

Examples infinite-dimensional vector spaces

- Space of all polynomials
- Space of all continuous functions on [***a***,***b***]
- Space of all continuous functions on [***a***,***b***] such that

$$f(a) = f(b) = 0$$

Vector norms

Definition 2. *Norm is a function defined on a vector space V:*

$$\mathcal{N} : V \longrightarrow \overline{\mathbb{R}}_+ \equiv [0, +\infty]$$

such that

$$(1)\ \ \|\mathbf{a}\| \geq 0, \quad \|\mathbf{a}\| = 0 \ \text{ iff } \ \mathbf{a} = \mathbf{0},$$

$$(2)\ \ \|\alpha\mathbf{a}\| = |\alpha|\|\mathbf{a}\|,$$

$$(3)\ \ \|\mathbf{a}+\mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|.$$

Examples

Example The space of continuous functions on the interval $[a, b]$ with the maximum norm

$$V = C([a, b]), \quad \|f\| = \sup_{[a,b]} |f(x)|.$$

If the interval is finite, $\|f\| = \max_{[a,b]} |f(x)|$.

Example The space of continuous functions on the interval $[a, b]$ with the maximum norm

$$V = L_p([a, b]), \quad \|f\| = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}.$$

Example The space $V = l_p$ of all sequences $\{a_k\}_{k=1}^{\infty}$ such that

$$\|\{a\}\| := \left(\sum_{k=1}^{\infty} |a_k|^p \right)^{1/p} < \infty.$$

In particular, l_1 is the space of all absolutely convergent sequences as

$$|a| := \sum_{k=1}^{\infty} |a_k| < \infty.$$

Example The space $V = l_\infty$ of all sequences $\{a_k\}_{k=1}^{\infty}$ such that

$$\|\{a\}\| := \sup_k |a_k| < \infty.$$

In other words, l_∞ is the space of all bounded sequences.

Inner product

Definition 3. *An inner product is a function* $(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ *or* $\mathbb{C}$ *satisfying*

$$(1)\ (\mathbf{a},\mathbf{a}) \geq 0, \quad (\mathbf{a},\mathbf{a}) = 0 \text{ iff } \mathbf{a} = \mathbf{0},$$

$$(2)\ (\mathbf{a},\mathbf{b}) = \overline{(\mathbf{b},\mathbf{a})},$$

$$(3)\ (\mathbf{a},\mathbf{b}+\mathbf{c}) = (\mathbf{a},\mathbf{b}) + (\mathbf{a},\mathbf{c}),$$

$$(4)\ (\alpha\mathbf{a},\mathbf{b}) = \alpha(\mathbf{a},\mathbf{b}).$$

Norm associated with an inner product is the 2-norm: $\| \boldsymbol{v} \| = (\boldsymbol{v}, \boldsymbol{v})^{1/2}$

Examples:

Legendre inner product $\quad f, g \in L_2([a,b]), \quad (f,g) = \int_a^b f(x)\overline{g(x)}dx$

Chebyshev inner product $\quad f, g \in C([-1,1]), \quad (f,g) = \int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}dx$

Hermite inner product $\quad f, g \in C([-\infty,\infty]), \quad (f,g) = \int_{-\infty}^{\infty} f(x)g(x)e^{-x^2}dx$

Matrix norm

Definition 4. *The norm of a matrix associated with the vector norm* $\|\cdot\|$ *is defined as*

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}. \tag{1}$$

The geometric sense of the matrix norm is the maximal elongation of a unit vector as a result of the corresponding linear transformation.

Exercise Let $A = (a_{ij})$ be an $m \times n$ matrix, $m \geq n$. Show that then:

(1) For the l_1-norm,

$$\|A\|_1 = \max_j \sum_i |a_{ij}|,$$

i.e., the maximal column sum of absolute values.

(2) For the max-norm or l_∞-norm

$$\|A\|_{\max} = \max_i \sum_j |a_{ij}|,$$

i.e., the maximal row sum of absolute values

Eigenvalues and eigenvectors

- **Diagonalizable** matrices

$$A = R\Lambda R^{-1} \equiv R\Lambda L$$

$$= \underbrace{\begin{bmatrix} r_1 & r_2 & \cdots & r_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}}_{\text{right eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{\text{eigenvalues}} \underbrace{\begin{bmatrix} l_1 & \rightarrow \\ l_2 & \rightarrow \\ \vdots & \\ l_n & \rightarrow \end{bmatrix}}_{\text{left eigenvectors}}$$

$$Ar_j = \lambda_j r_j$$

$$l_j A = \lambda_j l_j$$

- **Symmetric** matrices: ***eigenvalues are real***; there is an ***ONB consisting of eigenvectors***

- **Matrix 2-norm:** $\|A\| = \sqrt{\max_j \lambda_j(A^*A)}$

- **Defective** matrices and the **Jordan form**: $A = VJV^{-1}$

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix} \qquad J_k = \begin{bmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_k & 1 \\ & & & \lambda_k \end{bmatrix}$$

Example: $A = \begin{bmatrix} 1 & 10 \\ 0 & 1 \end{bmatrix}$

The Jordan form is rarely computed

In numerical linear algebra, the Jordan form is rarely computed. The reason is that it is unstable with respect to small perturbations of A. For example, consider a 16×16 matrix A

(5)
$$A := \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}.$$

It is already in the Jordan form consisting of a single block, and its unique eigenvalue of algebraic multiplicity 16 is zero. Indeed,

$$\det(\lambda I - A) = \lambda^{16} = 0.$$

Now consider a perturbation of A such that the zero at its bottom left corner is replaced with 10^{-16}:

(6)
$$A + \delta A := \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 10^{-16} & & & & 0 \end{bmatrix}.$$

The eigenvalues of $A + \delta A$ are the roots of

$$\det(\lambda I - A) = \lambda^{16} - 10^{-16} = 0.$$

There are 16 distinct complex eigenvalues located at the corners of the 16-gon in the complex plane:

$$\lambda_k = 0.1 e^{i2\pi k/16}, \quad k = 0, 1, \ldots 15.$$

Hence, the Jordan form of A will be $\mathrm{diag}\{\lambda_0, \ldots, \lambda_{15}\}$ which is not close to (6). Thus, we see that a perturbation of the size of the machine epsilon has a dramatic effect on the Jordan form and on the magnitudes of the eigenvalues of A.

The Schur form is often computed

$$A = QTQ^\top$$

$$T = \begin{bmatrix} \lambda_1 & t_{12} & t_{13} & \cdots & t_{1n} \\ & \lambda_2 & t_{23} & \cdots & t_{2n} \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & t_{n-1,n} \\ & & & & \lambda_n \end{bmatrix}$$

Q is orthogonal , i.e. $\mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = 1$

QR decomposition

Theorem 1. *Let A be $m \times n$, $m \geq n$. Suppose that A has full column rank. Then there exist a unique $m \times n$ orthogonal matrix Q, i.e., $Q^\top Q = I_{n\times n}$, and a unique $n \times n$ upper-triangular matrix R with positive diagonals $r_{ii} > 0$ such that $A = QR$.*

Proof. The proof of this theorem is given by the Gram-Schmidt orthogonalization process.

Algorithm 1: Gram-Schmidt orthogonalization

Input : matrix $A = [a_1\ a_2\ \dots\ a_n]$, $m \times n$, $\text{rank}(A) = n$.
Output: orthogonal matrix Q $m \times n$, $Q^\top Q = I_{n\times n}$, and upper-triangular $n \times n$ matrix R with $r_{ii} > 0$.

for $i = 1, \dots, N$ **do**
  $q_i = a_i$;
  for $j = 1, \dots, i-1$ **do**
    $\begin{cases} r_{ji} = q_j^\top a_i & \text{CGS} \\ r_{ji} = q_j^\top q_i & \text{MGS} \end{cases}$;
    $q_i = q_i - r_{ji} q_j$;
  end
  $r_{ii} = \|q_i\|$;
  $q_i = q_i / r_{ii}$;
end

Here CGS and MGS stand for the Classic Gram-Schmidt and the Modified Gram-Schmidt respectively. □

SVD

Theorem 2. [5] *Let A be an arbitrary $m \times n$ matrix with $m \geq n$. Then we can write*

$$A = U \Sigma V^\top,$$

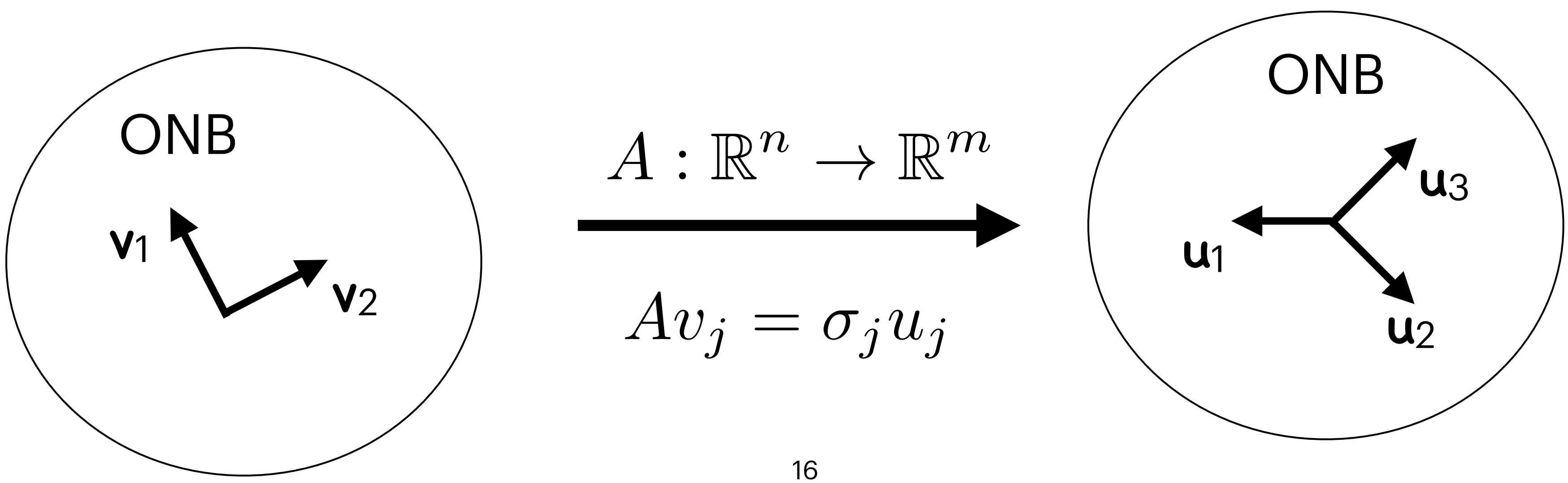
where

$$U \text{ is } m \times n \quad \text{and} \quad U^\top U = I_{n \times n},$$

$$\Sigma = \operatorname{diag}\{\sigma_1, \ldots, \sigma_n\}, \quad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0,$$

$$\text{and} \quad V \text{ is } n \times n \quad \text{and} \quad V^\top V = I_{n \times n}.$$

The columns of U, u_1, ..., u_n, are called left singular vectors. The columns of V, v_1, ..., v_n are called right singular vectors. The numbers σ_1, ..., σ_n are called singular values. If $m < n$, the SVD is defined for $A^\top$.



Theorem 3. *Let $A = U\Sigma V^\top$ be the SVD of the $m \times n$ matrix A, $m \geq n$.*

(1) *Suppose A is symmetric and $A = U\Lambda U^\top$ be an eigendecomposition of A Then the SVD of A is $U\Sigma V^\top$ where $\sigma_i = |\lambda_i|$ and $v_i = u_i \mathrm{sign}(\lambda_i)$, where $\mathrm{sign}(0) = 1$.*

(2) *The eigenvalues of the symmetric matrix $A^\top A$ are σ_i^2. The right singular vectors v_i are the corresponding orthonormal eigenvectors.*

(3) *The eigenvectors of the symmetric matrix $AA^\top$ are σ_i^2 and $m - n$ zeroes. The left singular vectors u_i are the corresponding orthonormal eigenvectors for the eigenvalues σ_i^2. One can take any $m-n$ orthogonal vectors as eigenvectors for the eigenvalue 0.*

(4) *If A has full rank, the solution of*

$$\min_x \|Ax - b\| \quad \text{is} \quad x = V\Sigma^{-1}U^\top b.$$

(5)

$$\|A\|_2 = \sigma_1.$$

If A is square and nonsingular , then

$$\|A^{-1}\|_2 = \frac{1}{\sigma_n}.$$

(6) *Suppose*

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0.$$

Then

$$\operatorname{rank}(A) = r,$$

$$\operatorname{null}(A) = \{x \in \mathbb{R}^n \ : \ Ax = 0 \in \mathbb{R}^m\} = \operatorname{span}(v_{r+1}, \ldots, v_n),$$

$$\operatorname{range}(A) = \operatorname{span}(u_1, \ldots, u_r).$$

(7)

$$A = U\Sigma V^\top = \sum_{i=1}^{n} \sigma_i u_i v_i^\top,$$

i.e., A is a sum of rank 1 matrices. Then a matrix of rank $k < n$ closest to A is

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top, \quad \text{and} \quad \|A - A_k\| = \sigma_{k+1}.$$

Condition number

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

The condition number is the ratio of the relative error in ***f*** to the relative error in ***x***:

$$\kappa(f;x) := \lim_{\epsilon \to 0} \max_{\|\Delta x\| = \epsilon} \frac{\frac{\|f(x+\Delta x) - f(x)\|}{\|f(x)\|}}{\frac{\|\Delta x\|}{\|x\|}}$$

Relative error in ***f***

Relative error in ***x***

$$f(x + \Delta x) = f(x) + J(x)\Delta x + O(\|\Delta x\|^2)$$

$$J(x) = U\Sigma V^\top. \text{ If } \Delta x \parallel v_1, \text{ then } \|J(x)\Delta x\| = \sigma_1 \|\Delta x\| \equiv \|J(x)\| \|\Delta x\|$$

$$\kappa(f;x) = \frac{\|J(x)\| \|x\|}{\|f(x)\|}$$

Condition number for matrix-vector multiplication

$$f(x) := Ax$$

$$\kappa(A;x) = \frac{\|A\|\|x\|}{\|Ax\|}$$ ... is large if $$\|A\| \gg \frac{\|Ax\|}{\|x\|}$$

$$A = U\Sigma V^\top$$

The worst-case scenario: $x \parallel v_n$

Example:

$$A = \begin{bmatrix} 1000 & 0 \\ 0 & 10 \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$ Then $$Ax = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$$

$$\Delta x = \begin{bmatrix} \epsilon \\ 0 \end{bmatrix}. \quad \text{Then} \quad A(x + \Delta x) - Ax = A\Delta x = \begin{bmatrix} 1000\epsilon \\ 0 \end{bmatrix}$$

Check that $\kappa(A;x) = 100$

Condition number for solving *Ax* = *b*

$$f(b) = A^{-1}b$$

$$\kappa(A^{-1}; b) = \frac{\|A^{-1}\|\|b\|}{\|A^{-1}b\|} = \|A^{-1}\|\frac{\|Ax\|}{\|x\|}$$

$$A = U\Sigma V^\top \qquad A^{-1} = V\Sigma^{-1}U^\top \qquad \|A^{-1}\| = \frac{1}{\sigma_n}$$

The worst-case scenario: $x \parallel v_1, \quad \text{or} \quad b \parallel u_1$

Example:

$A = \begin{bmatrix} 1000 & 0 \\ 0 & 10 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$ Then $x = A^{-1}b = \begin{bmatrix} 10^{-3} \\ 0 \end{bmatrix}$

$$\Delta b = \begin{bmatrix} 0 \\ \epsilon \end{bmatrix}. \quad \text{Then} \quad A^{-1}(b + \Delta b) - x = A^{-1}\Delta b = \begin{bmatrix} 0 \\ 0.1\epsilon \end{bmatrix}$$

Check that $\kappa(A^{-1}; b) = 100$