Review of linear algebra

Vector norms in $\mathbb{R}^n$ The p-norm of a vector $x \in \mathbb{R}^n$ is defined for $p \geq 1$ as

$$\|x\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$$

Special cases: Euclidean norm, $p = 2$,

$$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} = \sqrt{x^T \cdot x}$$

$p = 1$: 1-norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

The ∞-norm is defined as the maximum of the absolute values of the elements of x,

$$\|x\|_\infty = \max_{i=1:n} |x_i|$$

Remark on notation: All vectors are in column format $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

inner product in $\mathbb{R}^n$ is

$$x^T \cdot y = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

Eigenvalues and eigenvectors

Definition Given a matrix $A \in \mathbb{R}^{n \times n}$, a scalar number λ (real or complex) is called eigenvalue of A if the is a vector v such that $v \neq 0_n$ (nontrivial vector) and

$$\boxed{Av = \lambda v}$$

The vector v is called eigenvector associated with the eigenvalue λ.

Eigenvalues are solutions to the characteristic equation $\det(A - \lambda I) = 0$ where $I \in \mathbb{R}^{n \times n}$ denotes the identity matrix

Property Eigenvectors associated with distinct eigenvalues are linearly independent.

if the matrix A is symmetric, $A = A^T$, then eigenvectors associated with distinct eigenvalues are orthogonal,

$$\lambda_i \neq \lambda_j \Rightarrow v_i^T \cdot v_j = 0.$$

<u>Definition</u> A matrix $Q \in \mathbb{R}^{n \times n}$ is called <u>orthogonal</u> if the columns of Q are orthonormal vectors in $\mathbb{R}^n$:

$$\langle Q(:,i), Q(:,j) \rangle = \sum_{k=1}^{n} Q_{ki} Q_{kj} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

<u>Property</u> if Q is an orthogonal matrix then $Q^T Q = Q Q^T = I$

thus $Q^{-1} = Q^T$

For any two vectors $x, y \in \mathbb{R}^n$

$$x^T y = x^T Q^T Q y = (Qx)^T Q y$$

in particular, $\|Qx\|_2 = \|x\|_2$, $\forall x \in \mathbb{R}^n$

<u>Theorem</u> if $A \in \mathbb{R}^{n \times n}$ is a real and symmetric matrix then A can be written as

$$A = Q \Lambda Q^{-1} = Q \Lambda Q^T$$

where $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix consisting of eigenvectors of A, $A(:,i) = v_i$ and $\Lambda \in \mathbb{R}^{n \times n}$ is a <u>diagonal matrix</u> with diagonal entries the eigenvalues of A,

$$\Lambda_{ii} = \lambda_i$$

<u>Definition</u> A symmetric matrix is called <u>positive semidefinite</u> if for any vector $x \in \mathbb{R}^n$, $x^T A x \geq 0$.
The matrix is called <u>positive definite</u> if $x^T A x > 0$, $\forall x \in \mathbb{R}^n$, $x \neq 0_n$.

<u>Theorem</u> A symmetric matrix is positive semidefinite if and only if all its eigenvalues are greater than or equal to 0, $\lambda \geq 0$.
A symmetric matrix is positive definite if and only if all its eigenvalues are strictly positive, $\lambda > 0$.

<u>Remark</u> in Matlab the function eig(A) evaluates the eigenvalues and eigenvectors of a matrix A.

Matrix norms induced by vector norms

For a matrix $A \in \mathbb{R}^{n\times n}$ we define the matrix norm induced by the vector norm $\|\cdot\|$ in $\mathbb{R}^n$ as

$$\|A\| = \sup_{\|x\|=1} \|Ax\| = \max_{\|x\|=1} \|Ax\|$$

in particular, the p-norm of a matrix is defined as

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$$

The most popular matrix norms are:

2-norm: $\|A\|_2 = \sqrt{\lambda_{max}(A^TA)}$

where λ_{max} denotes the largest eigenvalue of A^TA.

1-norm:
(row norm) $\|A\|_1 = \max_{j=1:n} \sum_{i=1}^{n} |A_{ij}|$

∞-norm:
(column norm) $\|A\|_\infty = \max_{i=1:n} \sum_{j=1}^{n} |A_{ij}|$

May be extended to rectangular matrices.

<u>Definition</u> A matrix norm and a vector norm are compatible if

$$\|Ax\| \le \|A\| \cdot \|x\|$$

Notice that by definition the matrix p-norm is compatible with the vector p-norm.

<u>Definition</u> The <u>Frobenius norm</u> of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{trace}(A^T A)}$$

The Frobenius norm is compatible with the vector 2-norm (Euclidean norm)

$$\|Ax\|_2 \le \|A\|_F \|x\|_2$$

and is often used with the vector 2-norm

<u>Remark</u> Matlab command norm(A, p) has the options p = 1, 2, inf, 'fro' to evaluate the matrix norm of A.

Definition The condition number of a matrix A is defined as

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$$

Remark: condition number depends on the specified matrix norm. in Matlab, cond(A, p) can be used for p=1, 2, inf, 'fro'

The condition number of a matrix provides an upper bound on the error in the solution to the linear system

$$A x = b$$

in the presence of errors in the data vector b (measurement errors),

$$A \hat{x} = \hat{b}$$

We can bound the relative error in x $\frac{\|x - \hat{x}\|}{\|x\|}$ in terms of the relative error in data, $\frac{\|b - \hat{b}\|}{\|b\|}$

<u>Theorem</u> Let $A \in \mathbb{R}^{n \times n}$ nonsingular matrix.
Consider the linear systems

$$Ax = b \quad \text{and} \quad A\hat{x} = \hat{b}$$

Then $$\boxed{\frac{\|x-\hat{x}\|}{\|x\|} \le \|A\| \cdot \|A^{-1}\| \frac{\|b-\hat{b}\|}{\|b\|}}$$

where $\|A\|$ is a matrix norm compatible with the vector norm.

<u>Proof</u>: $A(x-\hat{x}) = b-\hat{b} \Rightarrow x-\hat{x} = A^{-1}(b-\hat{b})$

$\Rightarrow \|x-\hat{x}\| \le \|A^{-1}\| \, \|b-\hat{b}\|$ % absolute error bound

$$\Rightarrow \frac{\|x-\hat{x}\|}{\|Ax\|} \le \|A^{-1}\| \frac{\|b-\hat{b}\|}{\|b\|}$$

$$\Rightarrow \|x-\hat{x}\| \le \|A^{-1}\| \, \|Ax\| \frac{\|b-\hat{b}\|}{\|b\|} \le \|A\| \|A^{-1}\| \|x\| \frac{\|b-\hat{b}\|}{\|b\|}$$

$$\Rightarrow \frac{\|x-\hat{x}\|}{\|x\|} \le \|A\| \|A^{-1}\| \frac{\|b-\hat{b}\|}{\|b\|}$$

<u>Remark</u> if A is a symmetric and nonsingular matrix then

$$\mathrm{cond}(A)_2 = \frac{|\lambda_{max}|}{|\lambda_{min}|}$$

$|\lambda_{max}|$ → largest magnitude algebraic eig.

$|\lambda_{min}|$ → lowest magnitude algebraic eig

The singular value decomposition (SVD)

Let $A \in \mathbb{R}^{m \times n}$ a rectangular matrix with $m \geq n$

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ denote the eigenvalues of the matrix $A^T A$,

$$A^T A v_i = \lambda_i v_i , \quad i = 1:n \qquad (*)$$

The singular value decomposition (SVD) of A is a decomposition of the form

$$A = U S V^T$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, $UU^T = I_{m \times m}$, $VV^T = I_{n \times n}$ and $S \in \mathbb{R}^{m \times n}$ is a matrix such that $S_{ii} = \sqrt{\lambda_i} \overset{\text{notation}}{=} \sigma_i$ for $i = 1:n$ and zero entries elsewhere.

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ are called the <u>singular values</u> of A

Notice that $\sigma_i^2 = \lambda_i$ where λ_i are the eigenvalues of $A^T A$ $(*)$

The columns of the matrix U, $u_i \in \mathbb{R}^m$
$U = [u_1 \ldots\ldots u_m]$ are called the
left singular vectors of A

The columns of the matrix V, $v_i \in \mathbb{R}^n$,
$V = [v_1 \ldots\ldots v_n]$ are called the
right singular vectors of A

Notice that the right singular vectors are the eigenvectors of $A^T A \in \mathbb{R}^{n \times n}$:

$$A^T A v_i = \sigma_i^2 v_i, \quad i = 1:n$$

The left singular vectors are the eigenvectors of $AA^T \in \mathbb{R}^{m \times m}$,

$$AA^T u_i = \sigma_i^2 u_i, \quad i = 1:n$$

$$AA^T u_i = 0 \cdot u_i, \quad i = n+1:m$$

The singular values of A are unique.
The left and right singular vectors associated with singular values of multiplicity 1 are uniquely determined up to simultaneous sign change.

<u>Property</u>: The largest singular value of a matrix A is equal to its induced 2-norm,

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

The condition number of A measured in the 2-norm is

$$\text{cond}(A)_2 = \frac{\sigma_1}{\sigma_n}$$

<u>Additional properties of the SVD</u>

1) The Frobenius norm of A is

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$$

2) <u>Dyadic decomposition</u>: rank 1 outer products

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T$$

3) if rank $A = k$ then $\sigma_k > 0$ and $\sigma_{k+1} = 0$.

<u>Remark</u> The thin (short) form of the SVD for a matrix A or rank k is

$$A = USV^T, \quad U \in \mathbb{R}^{m\times k}, \quad S \in \mathbb{R}^{k\times k}, \quad V \in \mathbb{R}^{n\times k}$$

in Matlab, $[U, S, V] = \text{svd}(A)$
computes the SVD of the matrix A.
$\text{svds}(A, k)$ computes a short SVD
consisting of the first (leading) k
singular values and singular vectors.

Optimal approximation in the matrix 2-norm

The SVD provides the solution to the low
rank matrix approximation problem:

Given $A \in \mathbb{R}^{m \times n}$, $\text{rank}\, A = k \leq n \leq m$,
find matrix $X \in \mathbb{R}^{m \times n}$ of rank $\ell < k$
such that the error matrix $E = A - X$
is minimized:

$$\min_{X,\ \text{rank}(X) = \ell} \|A - X\|_2$$

<u>Lemma</u> Given $A \in \mathbb{R}^{m \times n}$ of rank k, for all $X \in \mathbb{R}^{m \times n}$ of rank $l \leq k$ we have

$$\|A - X\|_2 \geq \sigma_{l+1}(A)$$

<u>Theorem</u> (Schmidt - Eckart - Young - Mirsky)

$$\min_{X,\ \mathrm{rank}(X) = l} \|A - X\|_2 = \sigma_{l+1}(A)$$

provided that $\sigma_l > \sigma_{l+1}$. A minimizer X^* is obtained by truncating the dyadic decomposition to the first l terms

$$X^* = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_l u_l v_l^T$$

<u>Optimal approximation in Frobenius norm</u>

X^* defined above also provides an optimal approximation in the Frobenius norm. In addition,

$$\min_{X,\ \mathrm{rank}(X) = l} \|A - X\|_F = \left[\sum_{i=l+1}^{n} \sigma_i^2(A)\right]^{\frac{1}{2}}$$

<u>Remark</u> Notice that from the SVD

$A = USV^T$ it follows that

$$A v_i = \sigma_i u_i \,, \quad i = 1:n$$

$$A^T u_i = \sigma_i v_i \,, \quad i = 1:n$$

and $\quad A^T = V S^T U^T = \sum_{i=1}^{n} \sigma_i v_i u_i^T$

if $x \in \mathbb{R}^n$ is an arbitrary vector then

$$A x = \sum_{i=1}^{n} \sigma_i \, (v_i^T \cdot x) \, u_i$$

thus in the matrix vector multiplication the high-frequency components of x $(v_i^T \cdot x)$ are more damped as compared with the low-frequency components.

For ill-posed problems the condition number of A is large such that

$$\sigma_n \ll \sigma_1 \,.$$

In particular, if $A \in \mathbb{R}^{n \times n}$ is a nonsingular matrix then the solution to the linear system

$$Ax = b$$

is expressed as

$$x = \sum_{i=1}^{n} \left(\frac{u_i^T \cdot b}{\sigma_i} \right) v_i$$

if the vector b is corrupted by noise,

$$\hat{b} = b + \xi$$

the error in the solution x is

$$\hat{x} - x = \sum_{i=1}^{n} \left(\frac{u_i^T \cdot \xi}{\sigma_i} \right) v_i$$

the error term along the direction v_i is given by the noise component on direction u_i divided by the singular value σ_i :

$$\frac{u_i^T \cdot \xi}{\sigma_i}$$