0 Fundamental concepts and notation

This is a short summary of certain important fundamental techniques, which are usually covered in basic linear algebra courses.

0.1 Span and rank

We will need the basic concepts of the basis of vector space. Let $x_1, \ldots, x_k \in \mathbb{C}^n$, then we done the space these vectors span by

$$\operatorname{span}(x_1, \ldots, x_k) := \{c_1 x_1 + \cdots + c_k x_k : c_1, \ldots, c_k \in \mathbb{C}\} \tag{1}$$

or if we put the vectors into a matrix $X = [x_1, \ldots, x_k]$ the can be expressed with matrices

$$\operatorname{span}(x_1, \ldots, x_k) = \{Xc : c \in \mathbb{C}^k\} \tag{2}$$

also known as the range of the matrix X.

The rank of a matrix is the number of basis vectors needed to represent the columns. In other words the dimension of the column space.

Definition 0.1.1. *The matrix $A = [a_1, \ldots, a_m]$ has*

$$\operatorname{rank}(A) = k$$

where k is the smallest values such that there are vectors $v_1, \ldots, v_k$ which span the columns:

$$\operatorname{span}(a_1, \ldots, a_m) = \operatorname{span}(v_1, \ldots, v_k)$$

One of the most useful properties to analyze rank is in the context of matrix multiplication: product of two matrices is at most the rank of the individual matrices:

$$\operatorname{rank}(BC) \leq \min(\operatorname{rank}(B), \operatorname{rank}(C)). \tag{3}$$

Example: Rank 2

The matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

has rank 2. We have

$$\mathrm{span}(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix}) = \mathrm{span}(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix})$$

Transformation

If we denote $X = [x_1, \dots, x_k]$, the span is not modified if we multiply X from the right with a non-singular matrix. Let

$$Y = XF$$

where $F \in \mathbb{C}^{k\times k}$ is non-singular then the span remains the same

$$\mathrm{span}(x_1, \dots, x_k) = \mathrm{span}(y_1, \dots, y_k). \tag{4}$$

0.2 *Norms of vectors and matrices*

0.2.1 *Vector norms*

A norm is a quantification of the vector's magnitude. The most common vector norms for vector $v \in \mathbb{C}^k$ are

$$\begin{aligned} \|v\|_2 &= \sqrt{|v_1|^2 + \cdots + |v_k|^2} \\ \|v\|_\infty &= \max_{j=1,\dots,k} |v_k| \\ \|v\|_1 &= |v_1| + \cdots + |v_k| \end{aligned}$$

0.2.2 *Matrix norms*

Matrix magnitudes can be quantified similar to vectors. In the context of matrix norms we additionally require that it is submulticative

$$\|Av\| \le \|A\|\|v\|$$

The most common matrix norm is an *operator norm* which is defined from a vector norm by the relation

$$\|A\| := \sup_{\|v\|=1} \|Av\|.$$

The operator 1-norm and ∞-norm are easy to compute

$$\|A\|_1 := \sup_{\|v\|_1=1} \|Av\|_1 = \max_j \sum_i |a_{ij}| \tag{5}$$

$$\|A\|_\infty := \sup_{\|v\|_\infty=1} \|Av\|_\infty = \max_i \sum_j |a_{ij}|. \tag{6}$$

The operator two-norm, is computed from the largest eigenvalue of the matrix A^TA:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^TA)}$$

(which is the largest singular value, as we shall learn later.)

Apart from the norms above, the *Frobenius norm* is a common matrix norm

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2}.$$

0.3 Orthogonal matrices

Several of the matrix factorization that appear in this course involve orthogonal matrices. This is the formal definition of an orthogonal matrix.

Orthogonal matrices are important in matrix computations, most importantly when the matrix represents a basis of a vector space. Many properties of the vector space are not robust with respect to rounding errors if the basis is not orthogonal.

The identity matrix will be denoted $I \in \mathbb{R}^{k\times k}$, where k is given by the context we use it.

Definition 0.3.1 (Orthogonal matrix). *$Q \in \mathbb{R}^{n\times m}$ is called an orthogonal matrix if*

$$Q^TQ = I.$$

For complex matrices, the corresponding property is called unitary: $Q^*Q = I$.

Properties:

(i) The columns of an orthogonal matrix are orthonormal.

(ii) If $n = m$, then $Q^T = Q^{-1}$.

(iii) If $n = m$, then $QQ^T = I$.

Note that (iii) is not satisfied if Q is a rectangular matrix ($n \neq m$). For instance

$$Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$$

is an orthogonal matrix since $Q^TQ = I$, but

$$QQ^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

0.4 *Eigenvalue decomposition*

The eigenvalues λ_i and associated eigenvectors v_i of a square matrix $A \in \mathbb{C}^{n\times n}$ satisfy

$$Av_i = \lambda_i v_i.$$

In the standard situation (called diagonalizable matrices) we have n eigenvalues and n eigenvalues. For diagonalizable matrices we can factorize A into

$$A = V\Lambda V^{-1}$$

where $v = [v_1, \dots, v_n]$ and

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Not all matrices are diagonalizable. In this situation one needs to use theory for Jordan decompositions, which is not needed in this course.

Note that the eigenvalue decomposition can be used to determine the rank of square matrices. (A more robust approach is based on the singular value decomposition, which we will learn later.)

Detemining rank

If all eigenvalues (of a diagonalizable matrix) except one is zero: $\lambda_2 = \cdots = \lambda_n = 0$, then

$$A = [v_1, \dots, v_n] \begin{bmatrix} \lambda_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \begin{bmatrix} w_1^T \\ \vdots \\ w_n^T \end{bmatrix} \tag{7}$$

where we defined $w_1, \dots, w_n$ such that

$$V^{-1} = \begin{bmatrix} w_1^T \\ \vdots \\ w_n^T \end{bmatrix}$$

By multiplying out equation (7) we get

$$A = \lambda_1 v_1 w_1^T.$$

Therefore, $\text{rank}(A) = 1$, since the columns of A are all multiplies of the vector v_1. If $\lambda_1 = 0$ the rank is zero.

○

Determining rank again

If we apply the transformation rul (4) to the eigenvalue decomposition (for a diagonalizable matrix) we see that

$$\text{span}(a_1, \dots, a_n) = \text{span}(V\Lambda V^{-1}) = \text{span}(V\Lambda)$$

Therefore,

$$\text{rank}(A) = \text{rank}(V\Lambda)$$

Since

$$V\Lambda = \left[\lambda_1 v_1, \vdots, \lambda_n v_n\right]$$

which are linearly independent for any nonzero eigenvalue. Hence,

$$\text{rank}(A) = \text{rank}(\Lambda) = \text{number of non-zero eigenvalues}$$