

## Linear Algebra refresher

In this lecture we will go through some of the key concepts of linear algebra and inverse problem theory that are required to develop the theories of the different machine learning algorithm presented in this course. This is not meant to be an exhaustive treatise and students are strongly advised to take the ErSE 213 - Inverse Problems prior to this course.

Three key mathematical objects arise in the study of linear algebra:

**Scalars:**  $a \in \mathbb{R}$ , a single number represented by a lower case italic letter;

**Vectors:**  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^N$ , ordered collection of  $N$  numbers represented by a lower case bold letter; it is sometimes useful to extract a subset of elements by defining a set  $\mathbb{S}$  and add it to as a superscript,  $\mathbf{x}_{\mathbb{S}}$ . As an example, given  $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]^T \in \mathbb{R}^6$  and  $\mathbb{S} = 1, 3, 5$  we can define the vector  $\mathbf{x}_{\mathbb{S}} = [x_1, x_3, x_5]$  and its complementary vector  $\mathbf{x}_{-\mathbb{S}} = [x_2, x_4, x_6]$

**Matrices:**  $\mathbf{X} \in \mathbb{R}^{[N \times M]}$ , two dimensional collection of numbers represented by an upper case bold letter where  $N$  and  $M$  are referred to as the height and width of the matrix. More specifically a matrix can be written as

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,M} \\ \dots & \dots & \dots \\ x_{N,1} & x_{N,2} & x_{N,M} \end{bmatrix}$$

A matrix can be indexed by rows  $\mathbf{X}_{i,:}$  (i-th row), by columns  $\mathbf{X}_{:,j}$  (j-th column), and by element  $\mathbf{X}_{i,j}$  (i-th row, j-th column). A number of useful operations that are commonly applied on vectors and matrices are now described:

- Transpose:  $\mathbf{Y} = \mathbf{X}^T$ , where  $Y_{i,j} = X_{j,i}$
- Matrix plus vector:  $\mathbf{Y}_{[N \times M]} = \mathbf{X}_{[N \times M]} + \mathbf{z}_{[1 \times M]}$ , where  $Y_{i,j} = X_{i,j} + z_j$  ( $\mathbf{z}$  is added to each row of the matrix  $\mathbf{X}$ )
- Matrix-vector product:  $\mathbf{y}_{[N \times 1]} = \mathbf{A}_{[N \times M]} \mathbf{x}_{[M \times 1]}$ , where  $y_i = \sum_{j=1}^M A_{i,j} x_j$
- Matrix-vector product:  $\mathbf{y}_{[N \times 1]} = \mathbf{A}_{[N \times M]} \mathbf{x}_{[M \times 1]}$ , where  $y_i = \sum_{j=1}^M A_{i,j} x_j$
- Matrix-matrix product:  $\mathbf{C}_{[N \times K]} = \mathbf{A}_{[N \times M]} \mathbf{B}_{[M \times K]}$ , where  $C_{i,k} = \sum_{j=1}^M A_{i,j} B_{j,k}$
- Hadamard product (i.e., element-wise product):  $\mathbf{C}_{[N \times M]} = \mathbf{A}_{[N \times M]} \odot \mathbf{B}_{[N \times M]}$ , where  $C_{i,j} = A_{i,j} B_{i,j}$
- Dot product:  $a = \mathbf{x}_{[N \times 1]}^T \mathbf{y}_{[N \times 1]} = \sum_{i=1}^N x_i y_i$
- Identity matrix:  $\mathbf{I}_N = \text{diag}\{\mathbf{1}_N\}$ . Based on its definition, we have that  $\mathbf{I}_N \mathbf{x} = \mathbf{x}$  and  $\mathbf{I}_N \mathbf{X} = \mathbf{X}$
- Inverse matrix: given  $\mathbf{y} = \mathbf{A} \mathbf{x}$ , the inverse matrix of  $\mathbf{A}$  is a matrix that satisfies the following equality  $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}_N$ . We can finally write  $\mathbf{x} = \mathbf{A}^{-1} \mathbf{y}$
- Orthogonal vectors and matrices: given two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , they are said to be orthogonal if  $\mathbf{y}^T \mathbf{x} = 0$ . Given two matrices  $\mathbf{X}$  and  $\mathbf{Y}$ , they are

said to be orthogonal if  $\mathbf{Y}^T \mathbf{X} = \mathbf{I}_N$ . Orthogonal matrices are especially interesting because their inverse is simply  $\mathbf{X}^{-1} = \mathbf{X}^T$

- Matrix decomposition: like any scalar number can be decomposed into a product of prime numbers, a matrix  $\mathbf{A}$  can also be decomposed into a combination of vectors (i.e., eigenvectors) and scalars (i.e., eigenvalues).  
- Eigendecomposition: real-valued, square, symmetric matrices can be written as  $\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T$  where  $\lambda_i$  and  $\mathbf{v}_i$  are the eigenvalues and eigenvectors of the matrix  $\mathbf{A}$ , respectively. Eigenvectors are placed along the columns of the matrix  $\mathbf{V}$ , which is an orthogonal matrix (i.e.,  $\mathbf{V}^T = \mathbf{V}^{-1}$ ). Eigenvalues are placed along the diagonal of the matrix  $\mathbf{\Lambda} = \text{diag}\{\lambda\}$  and tell us about the rank of the matrix,  $\text{rank}(\mathbf{A}) = \#\lambda \neq 0$ . A *full rank* matrix is matrix whose eigenvalues are all non-zero and can be inverted. In this case the inverse of  $\mathbf{A}$  is  $\mathbf{A}^{-1} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^T$  - Singular value decomposition (SVD): this is a more general decomposition which can be applied to real-valued, non-square, non-symmetric matrices. Singular vectors  $\mathbf{u}$  and  $\mathbf{v}$  and singular values  $\lambda$  generalized the concept of eigenvectors and eigenvalues. The matrix  $\mathbf{A}$  can be decomposed as  $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T$  where  $\mathbf{D} = \mathbf{\Lambda}$  for square matrices,  $\mathbf{D} = [\mathbf{\Lambda} \mathbf{0}]^T$  for  $N > M$  and  $\mathbf{D} = [\mathbf{\Lambda} \mathbf{0}]$  for  $M > N$ . Similar to the eigendecomposition, in this case the inverse of  $\mathbf{A}$  is  $\mathbf{A}^{-1} = \mathbf{V} \mathbf{D}^{-1} \mathbf{U}^T$
- Conditioning: in general, it refers to how fast a function  $f(x)$  changes given a small change in its input  $x$ . Similarly for a matrix, conditioning is linked to the curvature of its associated quadratic form  $f(\mathbf{A}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$  and it generally indicates how rapidly this function changes as function of  $\mathbf{x}$ . It is defined as  $\text{cond}(\mathbf{A}) = \frac{|\lambda_{max}|}{|\lambda_{min}|}$ .

**Norms:** another important object that we will be using when defining cost functions for ML models are norms. A norm is a function that maps a vector  $\mathbf{x} \in \mathbb{R}^N$  to a scalar  $d \in \mathbb{R}$  and it can be loosely seen as measure of the length of the vector (i.e., distance from the origin). In general, the  $L^p$  norm is defined as:

$$\|\mathbf{x}\|_p = \left( \sum_i |x_i|^p \right)^{1/p} \quad p \geq 0$$

Popular norms are:

- Euclidean norm ( $L_2$ ):  $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ , is a real distance of a vector from the origin of the N-d Euclidean space. Note that  $\|\mathbf{x}\|_2^2 = \mathbf{x}^T \mathbf{x}$  and that  $\|\mathbf{x}\|_2 = 1$  for a unit vector;
- $L_1$  norm:  $\|\mathbf{x}\|_1 = \sum_i |x_i|$
- $L_0$  norm: number of non-zero elements in the vector  $\mathbf{x}$
- $L_\infty$  norm:  $\|\mathbf{x}\|_\infty = \max |x_i|$
- Frobenius norm (for matrices):  $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$ ,