May 2, 2014

Appendix: Brief Review of Vector Space

1 Introduction

A set of linear equations can be written in the form,

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \quad \cdots \quad + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \quad \cdots \quad + a_{2n}x_n &= b_2 \\
\cdots \\
a_{m1}x_1 + a_{m2}x_2 + \quad \cdots \quad + a_{mn}x_n &= b_m
\end{aligned}
\tag{1}
$$

where $x_j,\ j = 1, ..., n$ is a set of unkonws, $b_i,\ i = 1, ..., m$ are the right-hand side coefficients, and a_{ij} are the coefficients of the system. If $n = m$ this system of linear equations can be represented in a matricial form as

$$
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & \cdots\cdots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \cdots\cdots & a_{2n} \\
\cdots & \cdots & \cdots & \cdots\cdots & \cdots \\
a_{n1} & a_{n2} & a_{n3} & \cdots\cdots & a_{nn}
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{pmatrix}
\tag{2}
$$

or simply

$$
\underline{\underline{A}}\ \underline{x} = \underline{b} \tag{3}
$$

We can use the traditional notation a_{ij} to refer to the element in the i^{th} row and j^{th} column of matrix A.

2 A Few Useful Properties of Matrices

2.1 Shape of Matrices

Column Matrix: $\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}$

Row Matrix: $\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots a_{1n} \end{pmatrix}$

Null or Zero Matrix: $\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$

Identity Matrix: $I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$ or $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

Diagonal Matrix: $\begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix}$ or $a_{ij} = \begin{cases} d_{ii} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

Upper Triangular Matrix: $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$ or $a_{ij} = \begin{cases} a_{ij} & \text{if } i \leqslant j \\ 0 & \text{otherwise} \end{cases}$

Lower Triangular Matrix: $\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ or $a_{ij} = \begin{cases} a_{ij} & \text{if } i \geqslant j \\ 0 & \text{otherwise} \end{cases}$

Dense Matrix: $\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & 0 & a_{24} \\ a_{31} & 0 & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$

Sparse Matrix: $\begin{pmatrix} a_{11} & 0 & a_{13} & 0 \\ 0 & a_{22} & 0 & a_{24} \\ 0 & 0 & 0 & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix}$

Symmetric Matrix: $\underline{\underline{A}} = \underline{\underline{A}}^{\mathrm{T}}$, i.e., $a_{ij} = a_{ji} \ \ \forall i, j$, e.g., $A = \begin{pmatrix} 3 & 1 & 0 & 4 \\ 1 & 9 & 5 & 2 \\ 0 & 5 & 8 & 6 \\ 4 & 2 & 6 & 7 \end{pmatrix} = A^{\mathrm{T}}$

2.2 Invertible Matrix

An $n \times n$ square matrix A is called **invertible** (or nonsingular) if there exists a matrix A^{-1} such that,

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I$$

Not **all matrices have inverses –** this property is crucial when we are solving large linear systems – $Ax = b$. If we multiply the matricial equation by A^{-1},

$$\begin{aligned} &Ax = b \quad \times A^{-1} \\ &A^{-1}Ax = A^{-1}b \\ &Ix = A^{-1}b \Longrightarrow \quad x = A^{-1}b \end{aligned} \tag{4}$$

2.3 Operations with Invertible and Transposed Matrices

- $\left(A^T\right)^T = A$
- $\left(A^{-1}\right)^{-1} = A$
- $\left(A^{-1}\right)^T = \left(A^T\right)^{-1} = A^{-T}$
- If $A = BCD$, then $A^T = D^T C^T B^T$ and $A^{-1} = D^{-1}C^{-1}B^{-1}$
- $(A+B)^T = A^T + B^T$
- $(A+B)^{-1} \neq A^{-1} + B^{-1}$

3 Norms

Norms are mathematical entities used to measure the size of a vector or matrix. Norms can 'diagnose' which vector or matrix is *smaller* or *larger*.

3.1 Vector Norms

Let $\mathcal{S}$ be a vector space of finite-/infinite-dimension, and let $||\cdot||$ denote a mapping $\mathcal{S} \rightarrow \mathbb{R}$ with the following properties:

(i) For any nonzero **v**: $||v|| > 0$

(ii) For any scalar λ: $||\lambda v|| = |\lambda| ||v||$

(iii) For any two vectors **u** and **v**, the *triangle inequality* holds: $||u + v|| \leqslant ||u|| + ||v||$

Then $|| \cdot ||$ is called a **norm** for $\mathcal{S}$. An L_P norm, commonly denoted as $|| \cdot ||_p$, is defined for $p \geqslant 1$ as

$$||x||_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \tag{5}$$

The norm in Eqn. 5 is often called the *Minkowski or Hölder norm*. The L_p norm above satisfies all 3 conditions defined in (i-iii). The most common L_p norms for vectors are:

(a) *Manhattan norm* or ℓ_1: $||x||_1 = \sum_{i=1}^{n} |x_i|$

(b) *Euclidean norm* or ℓ_2: $||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$

(c) *Chebyshev norm* or *Infinity norm* or *Max norm* or ℓ_∞: $||x||_\infty = \lim_{p\to\infty} \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} = \max_{1\leqslant i \leqslant n} |x_i|$

Example: For,

(a) $\boldsymbol{u} = (2\ 3\ 6)^T \implies ||u||_1 = 11$, $||u||_2 = 7$ and $||u||_\infty = 6$

(b) $\boldsymbol{v} = (3\ 4\ 5)^T \implies ||v||_1 = 12$, $||v||_2 = 7.07$ and $||v||_\infty = 5$

3.2 Matrix Norms

The *matrix norm* is also represented by $|| \cdot ||$ and must satisfy the following properties (similar to *vector norms*),

(i) For any nonzero matrix $\boldsymbol{A}$: $||\boldsymbol{A}|| > 0$

(ii) For any scalar λ: $||\lambda \boldsymbol{A}|| = |\lambda| ||\boldsymbol{A}||$

(iii) For any two matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$,

- the *triangle inequality* holds: $||\boldsymbol{A} + \boldsymbol{B}|| \leqslant ||\boldsymbol{A}|| + ||\boldsymbol{B}||$
- the *consistency property* holds: $||\boldsymbol{AB}|| \leqslant ||\boldsymbol{A}|| ||\boldsymbol{B}||$

The *vector norms*, defined in Section 3.1, have the counterparts the *matrix norms* as,

(a) *1-norm* (i.e., maximum absolute column sum of the matrix): $||\boldsymbol{A}||_1 = \max\limits_{1\leqslant j\leqslant n} \sum\limits_{i=1}^{n} |a_{ij}|$

(b) *2-norm* or *Spectral norm*: $||\boldsymbol{A}||_2 = \begin{cases} \lambda_{\text{max}} & \text{if } \boldsymbol{A} = \boldsymbol{A}^T \\ \sigma_{\text{max}} & \text{if } \boldsymbol{A} \neq \boldsymbol{A}^T \end{cases}$. λ_{max} is the largest *eigenvalue* of $\boldsymbol{A}$ and $\sigma_{\text{max}} = \sqrt{\lambda_{\text{max}}\left(\boldsymbol{A}^T\boldsymbol{A}\right)}$

(c) *Infinity norm* (i.e., maximum absolute row sum of the matrix): $||\boldsymbol{A}||_\infty = \max\limits_{1\leqslant i\leqslant n} \sum\limits_{j=1}^{n} |a_{ij}|$

(d) *Frobenius norm* or *Hilbert-Schmidt norm*: $||\boldsymbol{A}||_F = \sqrt{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} |a_{ij}|^2}$

Example: For matrix $\boldsymbol{A} = \begin{pmatrix} 2 & -1 \\ 3 & 5 \end{pmatrix}$:

- $||\boldsymbol{A}||_1 = \max\left(|2| + |3|, |-1| + |5|\right) = \max\left(5, 6\right) = 6$
- $||\boldsymbol{A}||_\infty = \max\left(|2| + |-1|, |3| + |5|\right) = \max\left(3, 8\right) = 8$
- $||\boldsymbol{A}||_F = \sqrt{|2|^2 + |-1|^2 + |3|^2 + |5|^2} = \sqrt{39} = 6.245$
- $||\boldsymbol{A}||_2 = \max\left[\sqrt{\lambda\left(\boldsymbol{A}^T\boldsymbol{A}\right)}\right] = \max(2.228; 5.834) = 5.834$