Highlights of Linear Algebra

- Goal: understanding even more than solving

$$
\begin{cases}
\mathbf{Ax} = \mathbf{b} \rightarrow \text{Find } \mathbf{x} \left(\text{Is the vector b in the column space of } \mathbf{A}?\right) \\
\mathbf{Ax} = \boldsymbol{\lambda}\mathbf{x} \rightarrow \text{Find } \mathbf{x} \text{ and } \boldsymbol{\lambda} \begin{pmatrix} \text{eigenvector directions so that } \mathbf{Ax} \text{ keeps the same direction as } \mathbf{x} \\ \text{solve anything linear when we know every } \mathbf{x} \text{ and } \boldsymbol{\lambda} \end{pmatrix} \\
\mathbf{Av} = \boldsymbol{\sigma}\mathbf{u} \rightarrow \text{Find } \mathbf{v}, \mathbf{u} \text{ and } \boldsymbol{\sigma} \begin{pmatrix} \text{close but different, two vectors } \mathbf{u} \text{ and } \mathbf{v} \\ \mathbf{A}: \text{rectangular, full of data} \\ \text{what part of that data matrix is important?} \\ \text{Singular Value Decomposition : find its simplest pieces } \boldsymbol{\sigma}\mathbf{u}\mathbf{v}^T \\ \text{data science meets linear algebra in the SVD} \\ \text{Principal Component Analysis : find those pieces } \boldsymbol{\sigma}\mathbf{u}\mathbf{v}^T \end{pmatrix} \\
\left.\begin{array}{l} \text{Minimize } \dfrac{\|\mathbf{Ax}\|^2}{\|\mathbf{x}\|^2} \\ \text{Factor the matix } \mathbf{A} \\ \rightarrow \text{Factor } \mathbf{A} = (\text{columns}) \text{ times } (\text{rows}) \end{array}\right\} \rightarrow \begin{pmatrix} \text{singular vectors and} \\ \text{compute} \begin{cases} \text{the best } \hat{\mathbf{x}} \text{ in the least squares} \\ \text{principal comonent } \mathbf{v}_1 \text{ in PCA} \end{cases} \\ \text{fit the data} \end{pmatrix}
\end{cases}
$$

1. Multiplication Ax Using Columns of A

- Matrix-vector multiplication
  - Dot products: (row)·(column), computing, low level
  - Linear combination of the columns of A
- Combinations of the columns fill out the column space of A

$$A_1 = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 7 & 10 \end{bmatrix}, A_3 = \begin{bmatrix} 2 & 3 & 1 \\ 2 & 4 & 1 \\ 3 & 7 & 1 \end{bmatrix}$$

- Factor A into C times R
  - R: row-reduced echelon form of A

- Rank of A = rank of C: count independent columns
- = Dimension of the column space of A and C
  - Column rank = Row rank

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 5 \end{bmatrix}$$

2. Matrix-Matrix Multiplication AB

- AB = (m by n) times (n by p)
- Inner product: rows times columns
  - mp inner products, n multiplications each
- Outer product: columns times rows
  - n outer products, mp multiplications each
- AB = sum of rank one matrices

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix} = \text{rank one matrix}$$

$$\left(\mathbf{u}\mathbf{v}^T\right)^T = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 6 & 6 & 3 \\ 8 & 8 & 4 \\ 12 & 12 & 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} = \mathbf{v}\mathbf{u}^T$$

row space of $\mathbf{u}\mathbf{v}^T$ is the line through $\mathbf{v}$

Insight from Column times Row

- Why is the outer product approach essential in data science?
- We are looking for the important part of a matrix A
- We don't usually want the biggest number in A
- What we want more is the largest piece of A: those pieces are rank one matrices uv^T
- Factor A into CR

$$\begin{cases} \mathbf{A} = \mathbf{L}\mathbf{U} \\ \mathbf{A} = \mathbf{Q}\mathbf{R} \\ \mathbf{S} = \mathbf{Q}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{T} \\ \mathbf{A} = \mathbf{X}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1} \\ \mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{T} \end{cases}$$

3. Four Fundamental Subspaces

- Column space C(A)
- Row space C(A^T)
- Nullspace N(A)
- Left nullspace N(A^T)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \xrightarrow{m=2, n=2, r=1} \begin{cases} \mathbf{A}\mathbf{x} = \mathbf{0} \rightarrow \mathbf{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \rightarrow C\left(\mathbf{A}^T\right) \perp N(\mathbf{A}) \\ \mathbf{A}^T\mathbf{y} = \mathbf{0} \rightarrow \mathbf{y} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \rightarrow C(\mathbf{A}) \perp N\left(\mathbf{A}^T\right) \end{cases}$$

$$\mathbf{B} = \begin{bmatrix} 1 & -2 & -2 \\ 3 & -6 & -6 \end{bmatrix} \xrightarrow{m=2, n=3, r=1} \begin{Bmatrix} C(\mathbf{B}) \subset R^m \\ C\left(\mathbf{B}^T\right) \subset R^n \end{Bmatrix} \rightarrow \mathbf{B}\mathbf{x} = \mathbf{0} : (n - r) \text{ independent solutions}$$

The other two fundamental spaces come from the transpose matrix A^{T}. They are $\boldsymbol{N}(A^{\mathrm{T}})$ and $\boldsymbol{C}(A^{\mathrm{T}})$. We call $\boldsymbol{C}(A^{\mathrm{T}})$ the "row space of A" because the rows of A are the columns of A^{T}. What are those spaces for our 2 by 2 example?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad \text{transposes to} \quad A^{\mathrm{T}} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}.$$

Both columns of A^{T} are in the direction of $(1, 2)$. The line of all vectors $(c, 2c)$ is $\boldsymbol{C}(A^{\mathrm{T}})$ = row space of A. The nullspace of A^{T} is in the direction of $(3, -1)$:

Nullspace of A^{T} $\quad A^{\mathrm{T}}y = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ gives $\begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} 3c \\ -c \end{bmatrix}$.

The four subspaces $\boldsymbol{N}(A)$, $\boldsymbol{C}(A)$, $\boldsymbol{N}(A^{\mathrm{T}})$, $\boldsymbol{C}(A^{\mathrm{T}})$ combine beautifully into the big picture of linear algebra. Figure A2 shows how the nullspace $\boldsymbol{N}(A)$ is perpendicular to the row space $\boldsymbol{C}(A^{\mathrm{T}})$. Every input vector x splits into a row space part x_r and a nullspace part x_n. Multiplying by A always(!) produces a vector in the column space. Multiplication goes from left to right in the picture, from x to $Ax = b$.

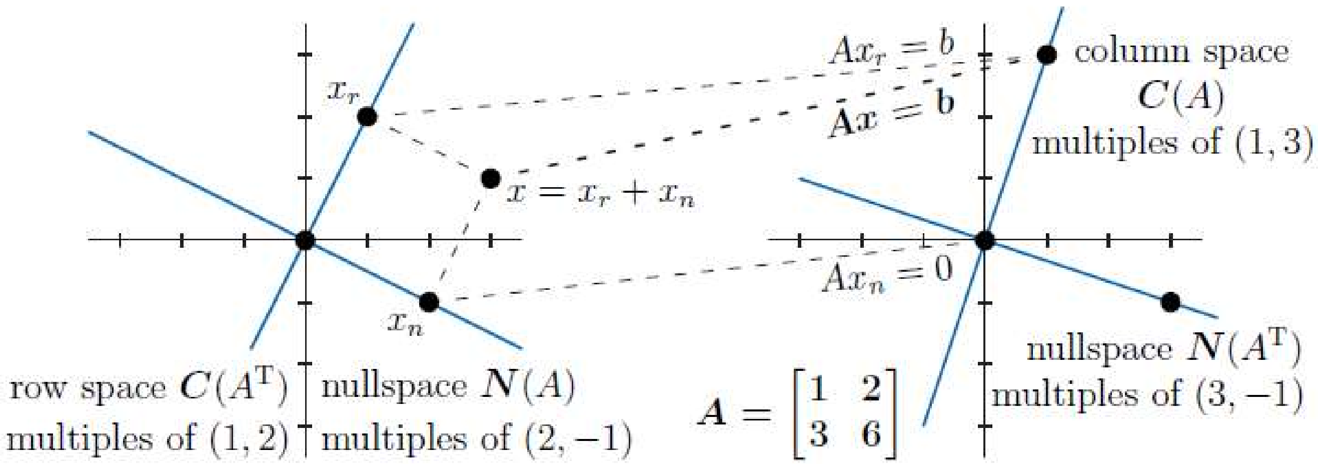

Figure A2: The four fundamental subspaces (lines) for the singular matrix A.

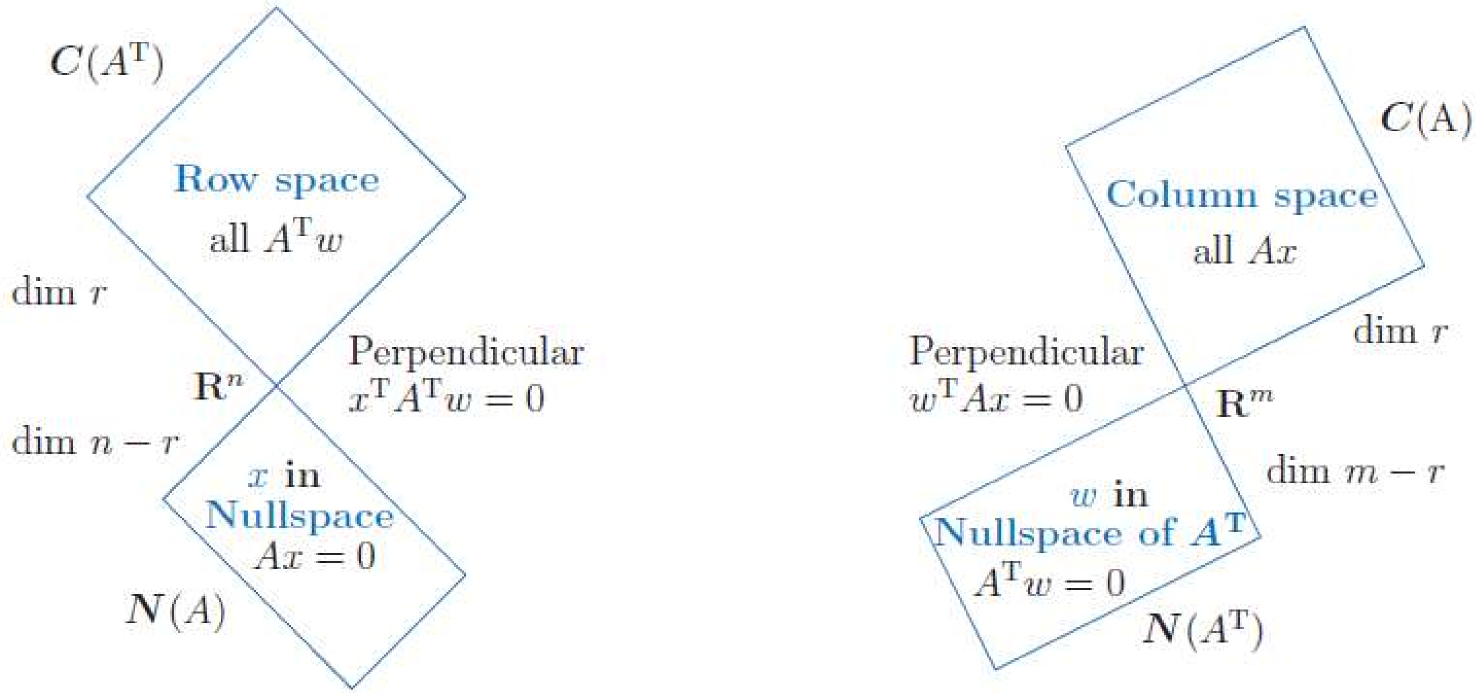

Figure A3: Dimensions and orthogonality for any m by n matrix A of rank r.

Figure A3 shows the Fundamental Theorem of Linear Algebra:

1. The row space in $\mathbf{R}^n$ and column space in $\mathbf{R}^m$ have the same dimension r.
2. The nullspaces $N(A)$ and $N(A^{\mathrm{T}})$ have dimensions $n - r$ and $m - r$.
3. $N(A)$ is perpendicular to the row space $C(A^{\mathrm{T}})$.
4. $N(A^{\mathrm{T}})$ is perpendicular to the column space $C(A)$.