A 2020 Vision of Linear Algebra

Gilbert Strang

MIT

2020

$A = \boldsymbol{CR} = \begin{bmatrix} & \\ & \\ & \end{bmatrix} \begin{bmatrix} & & \\ & & \end{bmatrix}$ **Independent columns in C**

$A = \boldsymbol{LU} = \begin{bmatrix} \diagdown & 0 \\ & \diagdown \end{bmatrix} \begin{bmatrix} \diagdown & \\ 0 & \diagdown \end{bmatrix}$ **Triangular matrices L and U**

$A = \boldsymbol{QR} = \begin{bmatrix} q_1 & q_n \end{bmatrix} \begin{bmatrix} \diagdown & \\ 0 & \diagdown \end{bmatrix}$ **Orthogonal columns in Q**

$S = \boldsymbol{Q\Lambda Q^{\mathrm{T}}} \quad Q^{\mathrm{T}} = Q^{-1}$ **Orthogonal eigenvectors $\boldsymbol{Sq = \lambda q}$**

$A = \boldsymbol{X \Lambda X^{-1}}$ Eigenvalues in Λ Eigenvectors in X $\boldsymbol{Ax = \lambda x}$

$A = \boldsymbol{U\Sigma V^{\mathrm{T}}}$ Diagonal $\Sigma =$ Singular values $\sigma = \sqrt{\boldsymbol{\lambda(A^{\mathrm{T}}A)}}$

Orthogonal vectors in $U^{\mathrm{T}}U = V^{\mathrm{T}}V = I$ $\boldsymbol{Av = \sigma u}$

$$A_0 = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 12 & 8 \\ 2 & 6 & 4 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 1 & 4 & 2 \\ 4 & 1 & 3 \\ 5 & 5 & 5 \end{bmatrix} \qquad S_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \qquad S_4 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

$$Q_5 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad A_6 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

Column space of A / All combinations of columns

$$
\begin{aligned}
A\boldsymbol{x} &= \begin{bmatrix} 1 & 4 & 5 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} x_3 \\
&= \text{linear combination of columns of} A
\end{aligned}
$$

Column space of A / All combinations of columns

$$Ax = \begin{bmatrix} 1 & 4 & 5 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} x_3$$

= linear combination of columns of A

Column space of $\boldsymbol{A}$ $= \mathbf{C}(A) =$ all vectors $A\boldsymbol{x}$

= all linear combinations of the columns

The column space of this example is

- R^3 ?
- plane ?
- line ?

Column space of A / All combinations of columns

$$Ax = \begin{bmatrix} 1 & 4 & 5 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} x_3$$

= linear combination of columns of A

Column space of A = $\mathbf{C}(A)$ = all vectors Ax

= all linear combinations of the columns

The column space of this example is
- R^3 ?
- plane ?
- line ?

Answer $\mathbf{C}(A)$ = **plane**

Basis for the column space / **Basis** for the row space

Include column $1 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ in C Include column $2 = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ in C

DO NOT INCLUDE COLUMN $3 = \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$
IT IS NOT INDEPENDENT

$$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R} = \begin{bmatrix} 1 & 4 \\ 3 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{bmatrix}$$

Row rank =
column rank =
$\boldsymbol{r} = \mathbf{2}$

The rows of R are a basis for the row space

$A = CR$ shows that column rank of A = row rank of A

1. The r columns of C are independent (by their construction)
2. Every column of A is a combination of those r columns (because $A = CR$)

3. The r rows of R are independent (they contain the r by r matrix I)
4. Every row of A is a combination of those r rows (because $A = CR$)

$A = CR$ shows that column rank of A = row rank of A

1. The r columns of C are independent (by their construction)
2. Every column of A is a combination of those r columns (because $A = CR$)
3. The r rows of R are independent (they contain the r by r matrix I)
4. Every row of A is a combination of those r rows (because $A = CR$)

Key facts

The r columns of C are a **basis** for the column space of A: **dimension** $\boldsymbol{r}$

The r rows of R are a **basis** for the row space of A: **dimension** $\boldsymbol{r}$

Basis for the column space / **Basis** for the row space

Include column $1 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ Include column $2 = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$

DO NOT INCLUDE COLUMN $3 = \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$
IT IS NOT INDEPENDENT

Basis has 2 vectors A has rank $r = 2$ $n - r = 3 - 2 = 1$

Counting Theorem $A\boldsymbol{x} = \mathbf{0}$ has one solution $\boldsymbol{x} = (1, 1, -1)$

There are $n - r$ independent solutions to $A\boldsymbol{x} = \mathbf{0}$

Matrix A with rank 1

If all columns of A are multiples of column 1,
show that all rows of A are multiples of one row

Proof using $A = CR$

One column $\boldsymbol{v}$ in $C \Rightarrow$ one row $\boldsymbol{w}$ in R

$$A = \begin{bmatrix} \\ \boldsymbol{v} \\ \\ \end{bmatrix} \begin{bmatrix} \ \boldsymbol{w} \ \end{bmatrix} \quad \Rightarrow \quad \textbf{all rows are multiples of } w$$

$A = CR$ is desirable $+$ $A = CR$ is undesirable $-$

C has columns directly from A : meaningful $+$

R turns out to be the **row reduced echelon form of A** $+$

Row rank $=$ Column rank is clear : $C =$ column basis, $R =$ row basis $+$

C and R could be very ill-conditioned $-$

If A is invertible then $C = A$ and $R = I$: **no progress** $\boldsymbol{A = AI}$ $-$

If $A\boldsymbol{x} = \mathbf{0}$ then $\begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ $\boldsymbol{x}$ is orthogonal to every row of A

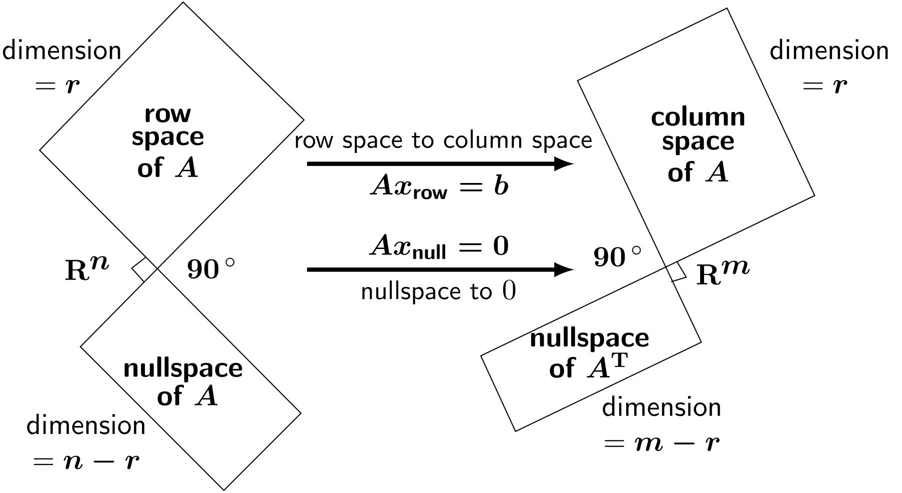
Every $\boldsymbol{x}$ in the nullspace of A is orthogonal to the row space of A

Every $\boldsymbol{y}$ in the nullspace of A^{T} is orthogonal to the column space of A

	N$(\boldsymbol{A})$ $\perp$	**C**$(\boldsymbol{A}^{\mathbf{T}})$	**N**$(\boldsymbol{A}^{\mathbf{T}})$ $\perp$	**C**$(\boldsymbol{A})$
Dimensions	$\boldsymbol{n-r}$	$\boldsymbol{r}$	$\boldsymbol{m-r}$	$\boldsymbol{r}$

Two pairs of **orthogonal subspaces**. The dimensions add to n and to m.

Big Picture of Linear Algebra



This is the Big Picture—two subspaces in $\mathbf{R}^n$ and two subspaces in $\mathbf{R}^m$.

From row space to column space, A is invertible.

$A = BC =$ sum of rank-1 matrices (**column times row** : **outer product**)

$$BC = \begin{bmatrix} | & | & & | \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdot\cdot & \boldsymbol{b}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & \boldsymbol{c}_1^* & - \\ - & \boldsymbol{c}_2^* & - \\ & \vdots & \\ - & \boldsymbol{c}_n^* & - \end{bmatrix} = \boldsymbol{b}_1 \boldsymbol{c}_1^* + \boldsymbol{b}_2 \boldsymbol{c}_2^* + \cdots + \boldsymbol{b}_n \boldsymbol{c}_n^*$$

New way to multiply matrices ! High level ! Row-column is low level !

$$\boldsymbol{A = LU} \quad \boldsymbol{A = QR} \quad \boldsymbol{S = Q\Lambda Q^{\mathrm{T}}} \quad \boldsymbol{A = X\Lambda X^{-1}} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}} \quad \boldsymbol{A = CR}$$

Elimination on $A\boldsymbol{x} = \boldsymbol{b}$ Triangular L and U

$$\begin{aligned} 2x + 3y &= 7 \\ 4x + 7y &= 15 \end{aligned} \qquad \begin{aligned} 2x + 3y &= 7 \\ y &= 1 \end{aligned} \qquad \begin{aligned} x &= 2 \\ y &= 1 \end{aligned}$$

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = LU$$

If rows are exchanged then $PA = LU$: **permutation $\boldsymbol{P}$**

Solve $Ax = b$ by elimination: **Factor $A = LU$**

Lower triangular L times upper triangular U

Step 1 Subtract ℓ_{i1} times row 1 from row i to produce zeros in column 1

$$\textit{Result}\quad A = \begin{bmatrix} 1 \\ \ell_{21} \\ \cdot \\ \ell_{n1} \end{bmatrix} \begin{bmatrix} \text{row 1 of } A \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & A_2 & \\ 0 & & & \end{bmatrix}$$

Step 2 Repeat Step 1 for A_2 then A_3 then $A_4 \ldots$

Step n L is lower triangular and U is upper triangular

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \ell_{21} & 1 & & \\ \cdot & \cdot & 1 & 0 \\ \ell_{n1} & \ell_{n2} & \ell_{n3} & 1 \end{bmatrix} \begin{bmatrix} \text{row 1 of } A \\ 0 \quad \text{row 1 of } A_2 \\ 0 \quad 0 \quad \text{row 1 of } A_3 \\ 0 \quad 0 \quad 0 \quad \text{row 1 of } A_n \end{bmatrix}$$

Orthogonal Vectors – Matrices – Subspaces

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0 \qquad \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0 \qquad (\boldsymbol{x}+\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}$ RIGHT TRIANGLE

Orthonormal columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ of Q: Orthogonal unit vectors

$$Q^{\mathrm{T}}Q = \begin{bmatrix} — & \boldsymbol{q}_1^{\mathrm{T}} & — \\ & \vdots & \\ — & \boldsymbol{q}_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} & & \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ & & \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} = I_n$$

$$QQ^{\mathrm{T}} = \begin{bmatrix} & & \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ & & \end{bmatrix} \begin{bmatrix} — & \boldsymbol{q}_1^{\mathrm{T}} & — \\ & \vdots & \\ — & \boldsymbol{q}_n^{\mathrm{T}} & — \end{bmatrix} = \boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}} + \cdots + \boldsymbol{q}_n\boldsymbol{q}_n^{\mathrm{T}} = \boldsymbol{I}$$

Orthogonal Vectors – Matrices – Subspaces

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0 \qquad \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0 \qquad (\boldsymbol{x}+\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}$ RIGHT TRIANGLE

Orthonormal columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ of Q: Orthogonal unit vectors

$$Q^{\mathrm{T}}Q = \begin{bmatrix} — & \boldsymbol{q}_1^{\mathrm{T}} & — \\ & \vdots & \\ — & \boldsymbol{q}_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} & & \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ & & \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} = I_n$$

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix} \qquad Q^{\mathrm{T}}Q = I \quad \boxed{QQ^{\mathrm{T}} \neq I} \qquad QQ^{\mathrm{T}}QQ^{\mathrm{T}} = QQ^{\mathrm{T}}$$

projection

"Orthogonal matrix"

$\boldsymbol{Q} = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$ **is square**. Then $QQ^{\mathrm{T}} = I$ and $Q^{\mathrm{T}} = Q^{-1}$

If Q_1, Q_2 are orthogonal matrices, so are Q_1Q_2 and Q_2Q_1

"Orthogonal matrix"

$\boldsymbol{Q} = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$ **is square**. Then $QQ^{\mathrm{T}} = I$ and $Q^{\mathrm{T}} = Q^{-1}$

If Q_1, Q_2 are orthogonal matrices, so are $Q_1 Q_2$ and $Q_2 Q_1$

$||Q\boldsymbol{x}||^2 = \boldsymbol{x}^{\mathrm{T}} Q^{\mathrm{T}} Q \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} = ||\boldsymbol{x}||^2$ Length is preserved

Eigenvalues of Q $\quad Q\boldsymbol{x} = \lambda \boldsymbol{x}$ $\quad ||Q\boldsymbol{x}||^2 = |\lambda|^2 \, ||\boldsymbol{x}||^2$ $\quad \boxed{|\lambda|^2 = 1}$

Rotation $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ $\quad \begin{aligned} \lambda_1 &= \cos\theta + i \sin\theta \\ \lambda_2 &= \cos\theta - i \sin\theta \end{aligned}$ $\quad |\lambda_1|^2 = |\lambda_2|^2 = 1$

Gram-Schmidt Orthogonalize the columns of A

$$\begin{aligned} A &= QR \\ Q^{\mathrm{T}}A &= R \\ \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{a}_k &= r_{ik} \end{aligned} \quad \begin{bmatrix} & & \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ & & \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \\ & r_{22} & \cdot & r_{2n} \\ & & \cdot & \cdot \\ & & & r_{nn} \end{bmatrix}$$

Columns $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ are **independent** Columns $\boldsymbol{q}_1$ to $\boldsymbol{q}_n$ are **orthonormal**!

Gram-Schmidt Orthogonalize the columns of A

$$\begin{array}{c} A = QR \\ Q^{\mathrm{T}}A = R \\ \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{a}_k = r_{ik} \end{array} \quad \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \\ & r_{22} & \cdot & r_{2n} \\ & & \cdot & \cdot \\ & & & r_{nn} \end{bmatrix}$$

Columns $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ are **independent** Columns $\boldsymbol{q}_1$ to $\boldsymbol{q}_n$ are **orthonormal**!

Column 1 of Q $\quad \boldsymbol{a}_1 = \boldsymbol{q}_1 r_{11} \qquad r_{11} = ||\boldsymbol{a}_1|| \qquad \boldsymbol{q}_1 = \dfrac{\boldsymbol{a}_1}{||\boldsymbol{a}_1||}$

Row 1 of $R = Q^{\mathrm{T}}A$ has $r_{1k} = \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_k$ Subtract (column) (row)

$$A - \boldsymbol{q}_1 \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_2 & \cdot & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} r_{22} & \cdot & r_{2n} \\ & \cdot & \cdot \\ & & r_{nn} \end{bmatrix}$$

Least Squares: Major Applications of $A = QR$

$\boldsymbol{m > n}$ m equations $A\boldsymbol{x} = \boldsymbol{b}$, n unknowns, minimize $||\boldsymbol{b} - A\boldsymbol{x}||^2 = ||\boldsymbol{e}||^2$

Least Squares: Major Applications of $A = QR$

$\boldsymbol{m} > \boldsymbol{n}$ m equations $A\boldsymbol{x} = \boldsymbol{b}$, n unknowns, minimize $||\boldsymbol{b} - A\boldsymbol{x}||^2 = ||\boldsymbol{e}||^2$

Normal equations for the best $\widehat{\boldsymbol{x}}$: $A^{\mathrm{T}}\boldsymbol{e} = \mathbf{0}$ or $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$

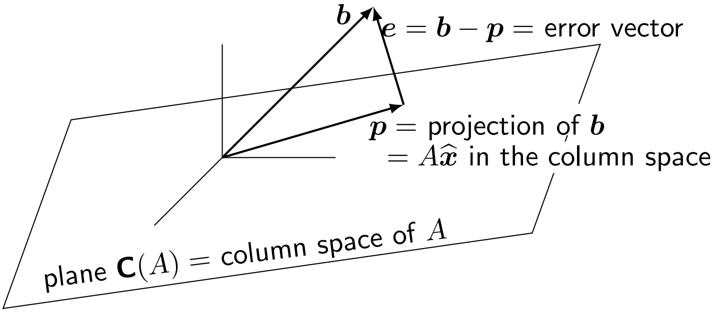
If $A = QR$ then $R^{\mathrm{T}}Q^{\mathrm{T}}QR\widehat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$ leads to $R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$

Least Squares: Major Applications of $A = QR$

$\boldsymbol{m > n}$ m equations $A\boldsymbol{x} = \boldsymbol{b}$, n unknowns, minimize $||\boldsymbol{b} - A\boldsymbol{x}||^2 = ||\boldsymbol{e}||^2$

Normal equations for the best $\widehat{\boldsymbol{x}}$: $A^{\mathrm{T}}\boldsymbol{e} = \mathbf{0}$ or $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$

If $A = QR$ then $R^{\mathrm{T}}Q^{\mathrm{T}}QR\widehat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$ leads to $R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$



$S = S^{\mathbf{T}}$ Real Eigenvalues and Orthogonal Eigenvectors

$S = S^{\mathrm{T}}$ has orthogonal eigenvectors $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. Important proof.

Start from these facts: $\boxed{S\boldsymbol{x} = \lambda\boldsymbol{x} \qquad S\boldsymbol{y} = \alpha\boldsymbol{y} \quad \lambda \neq \alpha \qquad S^{\mathrm{T}} = S}$

How to show orthogonality $\boldsymbol{x}^{\mathbf{T}}\boldsymbol{y} = \mathbf{0}$? Use every fact!

1. Transpose to $\boldsymbol{x}^{\mathrm{T}}S^{\mathrm{T}} = \lambda\boldsymbol{x}^{\mathrm{T}}$ and use $S^{\mathrm{T}} = S$ $\boxed{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{y} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}$
2. We can also multiply $S\boldsymbol{y} = \alpha\boldsymbol{y}$ by $\boldsymbol{x}^{\mathrm{T}}$ $\boxed{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{y} = \alpha\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}$
3. Now $\lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \alpha\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Since $\lambda \neq \alpha$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ **must be zero**

Eigenvectors of S go into Orthogonal Matrix Q

$$S\begin{bmatrix} \boldsymbol{q}_1 & \cdot\cdot & \boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{q}_1 & \cdot\cdot & \lambda_n\boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \cdot\cdot & \boldsymbol{q}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

That says $SQ = Q\Lambda$ $\qquad \boxed{\boldsymbol{S = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}}}$

$\boldsymbol{S = Q\Lambda Q^{\mathrm{T}}}$ is a sum $\lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}} + \cdots + \lambda_r\boldsymbol{q}_n\boldsymbol{q}_n^{\mathrm{T}}$ of rank one matrices

With $S = A^{\mathrm{T}}A$ this will lead to the singular values of A

$\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$ is a sum $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{T}}$ of rank one matrices

Singular values σ_1 to σ_r in Σ. Singular vectors in U and V

Eigenvalues and Eigenvectors of A: **Not symmetric**

$$A\left[\begin{array}{ccc} \boldsymbol{x}_1 & \cdot\cdot & \boldsymbol{x}_n \end{array}\right] = \left[\begin{array}{ccc} \lambda_1\boldsymbol{x}_1 & \cdot\cdot & \lambda_n\boldsymbol{x}_n \end{array}\right] \qquad \boldsymbol{AX = X\Lambda}$$

With n independent eigenvectors $\boldsymbol{A = X\Lambda X^{-1}}$

Eigenvalues and Eigenvectors of A: **Not symmetric**

$$A \left[\, \boldsymbol{x}_1 \;\; \cdot\cdot \;\; \boldsymbol{x}_n \,\right] = \left[\, \lambda_1 \boldsymbol{x}_1 \;\; \cdot\cdot \;\; \lambda_n \boldsymbol{x}_n \,\right] \qquad \boldsymbol{AX = X\Lambda}$$

With n independent eigenvectors $\boldsymbol{A = X\Lambda X^{-1}}$

$A^2, A^3, \ldots$ have the same eigenvectors as A

$$A^2\boldsymbol{x} = A(\lambda\boldsymbol{x}) = \lambda(A\boldsymbol{x}) = \lambda^2\boldsymbol{x} \qquad A^n\boldsymbol{x} = \lambda^n\boldsymbol{x}$$

$$A^2 = \left(X\Lambda X^{-1}\right)\left(XAX^{-1}\right) = X\Lambda^2X^{-1} \qquad \boldsymbol{A^n = X\Lambda^n X^{-1}}$$

$$A^n \to 0 \quad \text{when} \quad \Lambda^n \to 0 : \quad \textbf{All} \;\; |\boldsymbol{\lambda_i}| < \boldsymbol{1}$$

PROVE: $A^{\mathrm{T}}A$ is square, symmetric, nonnegative definite

1. $A^{\mathrm{T}}A = (n \times m)\,(m \times n) = n \times n$ Square

PROVE: $A^{\mathrm{T}}A$ is square, symmetric, nonnegative definite

1. $A^{\mathrm{T}}A = (n \times m)\,(m \times n) = n \times n$ Square

2. $(BA)^{\mathrm{T}} = A^{\mathrm{T}}B^{\mathrm{T}}$ $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}A^{\mathrm{TT}} = A^{\mathrm{T}}A$ Symmetric

PROVE: $A^{\mathrm{T}}A$ is square, symmetric, nonnegative definite

1. $A^{\mathrm{T}}A = (n \times m)\,(m \times n) = n \times n$ Square

2. $(BA)^{\mathrm{T}} = A^{\mathrm{T}}B^{\mathrm{T}}$ $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}A^{\mathrm{TT}} = A^{\mathrm{T}}A$ Symmetric

3. $S = S^{\mathrm{T}}$ is nonnegative definite IF

EIGENVALUE TEST $\mathbf{1}$: All eigenvalues ≥ 0 $S\boldsymbol{x} = \lambda\boldsymbol{x}$

ENERGY TEST $\mathbf{2}$: $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} \geq 0$ for every vector $\boldsymbol{x}$

PROVE: $A^{\mathrm{T}}A$ is square, symmetric, nonnegative definite

1. $A^{\mathrm{T}}A = (n \times m)\,(m \times n) = n \times n$ Square

2. $(BA)^{\mathrm{T}} = A^{\mathrm{T}}B^{\mathrm{T}}$ $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}A^{\mathrm{TT}} = A^{\mathrm{T}}A$ Symmetric

3. $S = S^{\mathrm{T}}$ is nonnegative definite IF

EIGENVALUE TEST **1**: All eigenvalues ≥ 0 $S\boldsymbol{x} = \lambda\boldsymbol{x}$

ENERGY TEST **2**: $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} \geq 0$ for every vector $\boldsymbol{x}$

TEST 1 IF $A^{\mathrm{T}}A\boldsymbol{x} = \lambda\boldsymbol{x}$ THEN $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ AND $\lambda = \dfrac{||A\boldsymbol{x}||^2}{||\boldsymbol{x}||^2} \geq 0$

TEST 2 applies to every $\boldsymbol{x}$, not only eigenvectors

Energy $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} = (A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x}) = ||A\boldsymbol{x}||^2 \geq 0$

Positive definite would have $\lambda > 0$ and $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} > 0$ for every $\boldsymbol{x} \neq 0$

AA^{T} is also symmetric positive semidefinite (or definite)

In applications $\dfrac{AA^{\mathrm{T}}}{n-1}$ can be the **sample covariance matrix**

$\boldsymbol{AA^{\mathrm{T}}}$ has the same nonzero eigenvalues as $\boldsymbol{A^{\mathrm{T}}A}$

Fundamental ! If $A^{\mathrm{T}}A\boldsymbol{x} = \lambda\boldsymbol{x}$ then $AA^{\mathrm{T}}A\boldsymbol{x} = \lambda A\boldsymbol{x}$

The eigenvector of $\boldsymbol{AA^{\mathrm{T}}}$ is $A\boldsymbol{x}$ ($\lambda \neq 0$ leads to $A\boldsymbol{x} \neq \mathbf{0}$)

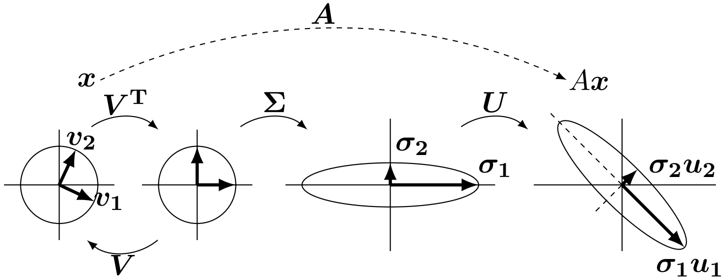
SINGULAR VALUE DECOMPOSITION

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} \text{ with } U^{\mathrm{T}}U = I \text{ and } V^{\mathrm{T}}V = I$$

$AV = U\Sigma$ means

$$A \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \text{ and } A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$$

SINGULAR VALUES $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ $\quad r =$ rank of A



U and V are rotations and possible reflections. Σ stretches circle to ellipse.

How to choose orthonormal $\boldsymbol{v}_i$ in the row space of A?

The $\boldsymbol{v}_i$ are eigenvectors of $A^{\mathrm{T}}A$

$A^{\mathrm{T}}A\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i$ The $\boldsymbol{v}_i$ are orthonormal. $\boldsymbol{V}^{\mathbf{T}}\boldsymbol{V} = \boldsymbol{I}$

How to choose orthonormal $\boldsymbol{v}_i$ in the row space of A?

The $\boldsymbol{v}_i$ are eigenvectors of $A^{\mathrm{T}}A$

$A^{\mathrm{T}}A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i$ The $\boldsymbol{v}_i$ are orthonormal. $\boldsymbol{V^{\mathrm{T}}V = I}$

How to choose $\boldsymbol{u}_i$ in the column space? $\boldsymbol{u}_i = \dfrac{A\boldsymbol{v}_i}{\sigma_i}$

The $\boldsymbol{u}_i$ are orthonormal This is the important step $U^{\mathrm{T}}U = I$

$$\left(\frac{A\boldsymbol{v}_j}{\sigma_j}\right)^{\mathrm{T}}\left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right) = \frac{\boldsymbol{v}_j^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{v}_i}{\sigma_j\sigma_i} = \frac{\boldsymbol{v}_j^{\mathrm{T}}\sigma_i^2\boldsymbol{v}_i}{\sigma_j\sigma_i} = \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix}$$

How to choose orthonormal $\boldsymbol{v}_i$ in the row space of A?

The $\boldsymbol{v}_i$ are eigenvectors of $A^{\mathrm{T}}A$

$A^{\mathrm{T}}A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i$ The $\boldsymbol{v}_i$ are orthonormal. $\boldsymbol{V^{\mathrm{T}}V = I}$

How to choose $\boldsymbol{u}_i$ in the column space? $\boldsymbol{u}_i = \dfrac{A\boldsymbol{v}_i}{\sigma_i}$

The $\boldsymbol{u}_i$ are orthonormal   This is the important step $U^{\mathrm{T}}U = I$

$$\left(\frac{A\boldsymbol{v}_j}{\sigma_j}\right)^{\mathrm{T}}\left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right) = \frac{\boldsymbol{v}_j^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{v}_i}{\sigma_j\sigma_i} = \frac{\boldsymbol{v}_j^{\mathrm{T}}\sigma_i^2\boldsymbol{v}_i}{\sigma_j\sigma_i} = \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix}$$

Full size SVD $A = U\Sigma V^{\mathrm{T}}$

$$m \times n \quad m \times m \quad n \times n$$

$\boldsymbol{u}_{r+1}$ to $\boldsymbol{u}_m$:   Nullspace of A^{T}

$\boldsymbol{v}_{r+1}$ to $\boldsymbol{v}_n$:   Nullspace of A

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & : & & \\ & & \sigma_r & \\ 0 & & & 0 \end{bmatrix}$$

SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ $\quad A^{\mathrm{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$ $\quad AA^{\mathrm{T}} = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$

$$\boldsymbol{U} = \frac{\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}}{\sqrt{10}} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 3\sqrt{5} & \\ & \sqrt{5} \end{bmatrix} \qquad \boldsymbol{V}^{\mathbf{T}} = \frac{\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}{\sqrt{2}}$$

$$\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

Low rank approximation to a big matrix

Start from the SVD $A = U\Sigma V^{\mathrm{T}} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$

Keep the k largest σ_1 to σ_k $A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}$

A_k is the closest rank k matrix to A $||\boldsymbol{A} - \boldsymbol{A_k}|| \leq ||\boldsymbol{A} - \boldsymbol{B_k}||$

Norms

$||A|| = \sigma_{\text{max}} \quad ||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \quad ||A||_N = \sigma_1 + \cdots + \sigma_r$

For very large matrices, randomization has brought a revolution

Example : Multiply AB with Column-row sampling $(AS)\,(S^{\mathrm{T}}B)$

$$AS = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \begin{bmatrix} s_{11} & 0 \\ 0 & 0 \\ 0 & s_{32} \end{bmatrix} = \begin{bmatrix} s_{11}\boldsymbol{a}_1 & s_{32}\boldsymbol{a}_3 \end{bmatrix} \text{ and } S^{\mathrm{T}}B = \begin{bmatrix} s_{11} & b_1^{\mathrm{T}} \\ s_{32} & b_3^{\mathrm{T}} \end{bmatrix}$$

NOTICE SS^{T} is not close to I. But we can have

$$\boldsymbol{E}[SS^{\mathrm{T}}] = I \qquad \boldsymbol{E}[(AS)\,(S^{\mathrm{T}}B)] = AB$$

Norm-squared sampling Choose column-row with probabilities $\approx ||a_i||\,||b_i^{\mathrm{T}}||$

This choice minimizes the **sampling variance**

OCW.MIT.EDU and YouTube

Math 18.06 Introduction to Linear Algebra

Math 18.065 Linear Algebra and Learning from Data

Math 18.06 Linear Algebra for Everyone (New textbook expected in 2021 !!)

math.mit.edu/linearalgebra **math.mit.edu/learningfromdata**

Resource: A 2020 Vision of Linear Algebra
Gilbert Strang