Key Ideas in Linear Algebra

Gilbert Strang

MIT

2019

Multiply $A\boldsymbol{x}$ by columns, not rows

$$A\boldsymbol{x} = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{a}_1 \\ \\ \end{bmatrix} x_1 + \cdots + \begin{bmatrix} \\ \boldsymbol{a}_n \\ \\ \end{bmatrix} x_n$$

$A\boldsymbol{x}$ is a *combination of the columns* $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$

Column space $\mathbf{C}(A) =$ **all** combinations of the columns

$A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution $\boldsymbol{x}$ when $\boldsymbol{b}$ is in $\mathbf{C}(A)$

Matrix multiplication : Columns times rows

$$AB = \begin{bmatrix} \boldsymbol{a}_1 \dots \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1^* \\ \vdots \\ \boldsymbol{b}_n^* \end{bmatrix} = \boldsymbol{a}_1\boldsymbol{b}_1^* + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^*$$

Sum of rank-one matrices $\boldsymbol{a}_i\boldsymbol{b}_i^* =$ **column times row**

$$\text{Example} \quad \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ 8 & 4 & -4 \\ 6 & 3 & -3 \end{bmatrix}$$

Column space = all multiples of $\begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$ Row space = all multiples of $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$

Dimension of column space $= 1 =$ Dimension of row space : Rank $= 1$

Basis for the column space – by example

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

Column 1 is not zero – it goes in the basis

Column 2 is not a multiple of column 1 – put into the basis

Column 3 is (Column 1) $-$ (Column 2): Dependent column

Column basis matrix $C = \begin{bmatrix} 6 & 2 \\ 4 & 2 \\ 3 & 2 \end{bmatrix}$

Column space has dimension 2 (2 vectors in the basis)

Row space has what dimension ??

Dimension of row space = Dimension of column space

Proof by factoring $A = CR$ to see **row rank = column rank**

$$\begin{bmatrix} 6 & 2 & 4 \\ 4 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 4 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$A \qquad\qquad C \qquad\qquad R$$

Columns of $A =$ combination of columns of C **Column basis in $\boldsymbol{C}$**

Rows of A $=$ combination of rows of $R =$ **Row basis in $\boldsymbol{R}$**

$$\boxed{\textbf{RANK of } \boldsymbol{A} \qquad \boldsymbol{r = 2 = 2}}$$

$\boldsymbol{A = CMR}$ has become an important factorization

Mixing matrix $\boldsymbol{M}$ $=$ invertible r by r **$\boldsymbol{C}$ and $\boldsymbol{R}$ come from $\boldsymbol{A}$**

Four great factorizations

1.	Symmetric $S = Q\Lambda Q^{\mathrm{T}}$	eigenvectors in Q
		eigenvalues in Λ
2.	Every $A = U\Sigma V^{\mathrm{T}}$	left singular vectors in U
		singular values in Σ
		right singular vectors in V
3.	Orthogonalize columns $A = QR$	Orthogonal Q
		Triangular R
4.	Elimination $A = LU$	Lower triangular L
	No row exchanges	Upper triangular U

Q has n orthonormal columns (length 1)

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

If Q is **square** then $\boldsymbol{Q}^{\mathbf{T}} = \boldsymbol{Q}^{-\mathbf{1}}$ "**orthogonal matrix**" $(QQ^{\mathrm{T}} = I)$

If Q is **rectangular** then QQ^{T} is a projection P $\quad (QQ^{\mathrm{T}} \neq I)$

$$P^2 = QQ^{\mathrm{T}}QQ^{\mathrm{T}} = QQ^{\mathrm{T}} = P$$

Orthogonal matrices (square) are great for computation

$\boxed{||Q\boldsymbol{x}|| = ||\boldsymbol{x}||}$ $\quad$ $\boxed{Q_1Q_2 \text{ is also orthogonal}}$

Symmetric $\boxed{\boldsymbol{S} = \boldsymbol{Q\Lambda Q}^{\mathbf{T}} = \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^{\mathrm{T}} + \cdots + \lambda_n \boldsymbol{q}_n \boldsymbol{q}_n^{\mathrm{T}}}$

Positive definite if all $\lambda_i > 0$

Positive semidefinite if all $\lambda_i \geq 0$

Positive definite energy $\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} > 0$ all $\boldsymbol{x} \neq \mathbf{0}$

Positive semidefinite energy $\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} \geq 0$ all $\boldsymbol{x}$

Positive definite factorization $S = A^{\mathrm{T}} A$ full rank A

Positive semidefinite factorization $S = A^{\mathrm{T}} A$ any rank A

Key $\quad \boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} A^{\mathrm{T}} A \boldsymbol{x} = (A\boldsymbol{x})^{\mathrm{T}} (A\boldsymbol{x}) \geq 0$

$A = U\Sigma V^{\mathrm{T}} =$ (orthogonal)(diagonal)(orthogonal) $=$ **SVD**

$$AV = A\begin{bmatrix} \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ \\ \end{bmatrix} = U\Sigma = \begin{bmatrix} \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ \\ \end{bmatrix}\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}$$

$$\boxed{A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1 \;\cdots\; A\boldsymbol{v}_r = \sigma_r\boldsymbol{u}_r \qquad A \text{ has rank } r}$$

$$A^{\mathrm{T}}A = (V\Sigma^{\mathrm{T}}U^{\mathrm{T}})\,(U\Sigma V^{\mathrm{T}}) = V(\Sigma^{\mathrm{T}}\Sigma)V^{\mathrm{T}}$$

$\boldsymbol{v}$'s are eigenvectors of $A^{\mathrm{T}}A$: orthonormal!

σ^2 are eigenvalues of $A^{\mathrm{T}}A$: $\sigma^2 = \lambda \geq 0$!

$\boldsymbol{u}$'s are eigenvectors of AA^{T}: orthonormal!

$$\boldsymbol{u}_i = \frac{A\boldsymbol{v}_i}{\sigma_i} \quad \text{leads to} \quad \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = \frac{\boldsymbol{v}_i^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{v}_j}{\sigma_i\,\sigma_j} = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j\frac{\sigma_j}{\sigma_i} = \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix}$$

$$A = U\Sigma V^{\mathrm{T}} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}} \quad (\text{decreasing } \ \sigma_i)$$

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

PCA $=$ Principal Component Analysis uses the SVD

$$\boxed{\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} = \frac{3}{2} \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} = \text{rank } 1 \text{ matrix closest to } A}$$

$$A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} = \ \text{rank } k \text{ matrix closest to } \ A$$

$$A_k \ \text{ minimizes } \ ||A - \text{any rank } k \text{ matrix } \ ||$$

$$\ell^2 \text{ norm } ||A|| = \text{max } ||A\boldsymbol{x}||/||\boldsymbol{x}|| = \sigma_1 = \text{ largest } \sigma$$

$$\text{Frobenius norm } ||A||_F^2 = \text{ sum of all } |a_{ij}|^2 = \text{ sum of all } \sigma_i^2$$

$A = QR =$ (orthogonal)(triangular) **Gram-Schmidt**

$\boldsymbol{a}_1 = \boldsymbol{q}_1 r_{11}$ First columns of A and Q: $r_{11} = ||\boldsymbol{a}_1||$

$\boldsymbol{a}_2 = \boldsymbol{q}_1 r_{12} + \boldsymbol{q}_2 r_{22}$ Second columns: $r_{12} = \boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{a}_2$ and $r_{22} = ||\boldsymbol{a}_2 - \boldsymbol{q}_1 r_{12}||$

Every $r_{ij} = \boldsymbol{q}_i^{\mathrm{T}} \boldsymbol{a}_j$ Subtract each $\boldsymbol{q}_i\, r_{ij}$ $(i < j)$ from later columns $\boldsymbol{a}_j$

Version $\mathbf{1}$: Subtract when you reach $\boldsymbol{a}_j$ in step j

Version $\mathbf{2}$: Subtract as soon as you know $\boldsymbol{q}_i$ in step i

$\#\mathbf{2}$ allows column permutations: choose largest column in next step $i+1$

Then columns are permuted and $AP = QR$ is numerically stable

$A = LU =$ (lower triangular with $\ell_{ii} = 1$)(upper triangular)

First row of U $\quad \boldsymbol{u}_1^* = \boldsymbol{a}_1^* =$ first row of A

First column of L $\quad \boldsymbol{\ell}_1 =$ (first column of A)$/a_{11}$

Remove $\boldsymbol{\ell}_1 \boldsymbol{u}_1^*$ to leave $A - \boldsymbol{\ell}_1 \boldsymbol{u}_1^* = \begin{bmatrix} 0 & \mathbf{0}^* \\ \mathbf{0} & \boldsymbol{A_1} \end{bmatrix}$ $\qquad A_1$ has size $n-1$

Remove $\boldsymbol{\ell}_k \boldsymbol{u}_k^*$ to find $A - \sum_1^k \boldsymbol{\ell}_i \, \boldsymbol{u}_i^* = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{A_k} \end{bmatrix}$ $\quad A_k$ has size $n-k$

Note that $\boldsymbol{\ell_k}$ and $\boldsymbol{u}_k^*$ start with $k-1$ zeros: L and U are triangular

Finally $A = \sum_1^n \boldsymbol{\ell}_i \, \boldsymbol{u}_i^* = LU =$ (lower triangular) (upper triangular)

Ordering each of the factorizations

$\boldsymbol{S} = \boldsymbol{Q\Lambda Q}^{\mathbf{T}}$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$

$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathbf{T}}$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$

$\boldsymbol{AP} = \boldsymbol{QR}$ with $r_{11} \geq r_{22} \geq \ldots =$ "**column pivoting**"

$\boldsymbol{PA} = \boldsymbol{LU}$ with all $|\ell_{ij}| \leq 1 =$ "**partial pivoting**"

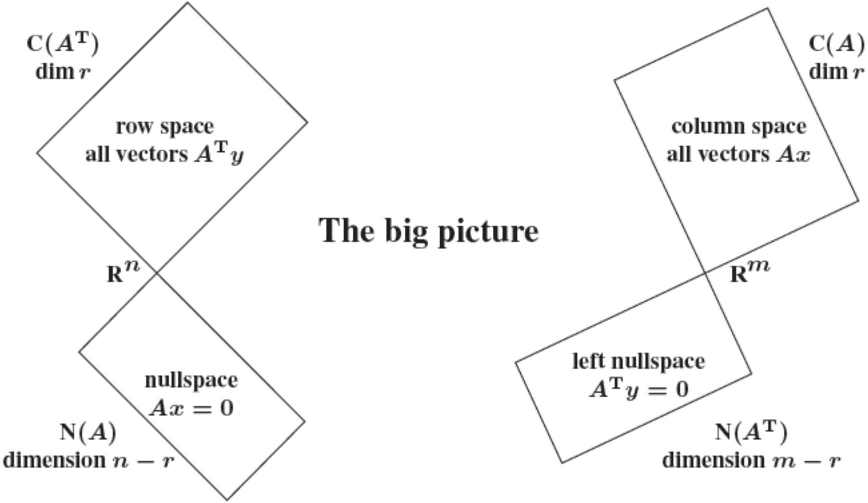
Those permutations P give numerical stability against roundo

Derivatives when $S = S(t)$ and $A = A(t)$

$$\frac{d\lambda_k(S)}{dt} = \boldsymbol{q}_k \frac{dS}{dt} \boldsymbol{q}_k^{\mathrm{T}}$$

$$\frac{d\sigma_k(A)}{dt} = \boldsymbol{u}_k \frac{dA}{dt} \boldsymbol{v}_k^{\mathrm{T}}$$

Four Fundamental Subspaces for A



Four Fundamental Subspaces: Their dimensions add to n and m

Pseudoinverse of $A = U\Sigma V^{\mathrm{T}}$ is $A^+ = V\Sigma^+ U^{\mathrm{T}}$

Pseudoinverse $\boldsymbol{\Sigma} \ = \text{diag}\,(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$ is m by n

$\boldsymbol{\Sigma}^+ = \text{diag}\,(1/\sigma_1, \ldots, 1/\sigma_r, 0, .., 0)$ is n by m

From row space to column space any A is invertible

From column space to row space A^+ is that inverse

$$\boxed{\begin{array}{l} \boldsymbol{A^+A} = \textbf{projection onto row space} \\ \boldsymbol{AA^+} = \textbf{projection onto column space} \end{array}}$$

$A^+\boldsymbol{b}$ is the minimum norm least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$

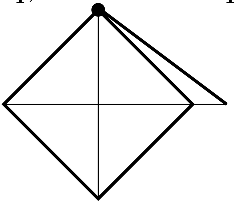
That is because $A^+\boldsymbol{b}$ has zero component in the nullspace of A

$A^+\boldsymbol{b}$ minimizes $||A\boldsymbol{x} - \boldsymbol{b}||^2 + \lambda||\boldsymbol{x}||^2$ as λ drops to zero

Minimization in ℓ^1 ℓ^2 ℓ^∞

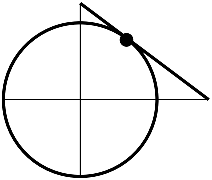
Minimize $||\boldsymbol{v}||$ among vectors (v_1, v_2) on the line $3v_1 + 4v_2 = 1$

$\left(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{4}}\right)$ has $||\boldsymbol{v}^*||_{\mathbf{1}} = \frac{\mathbf{1}}{\mathbf{4}}$



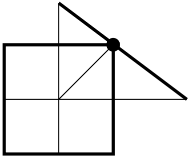
$\boldsymbol{\ell}^1$ diamond

$\left(\frac{\mathbf{3}}{\mathbf{25}}, \frac{\mathbf{4}}{\mathbf{25}}\right)$ has $||\boldsymbol{v}^*||_{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{5}}$



$\boldsymbol{\ell}^2$ circle

$\left(\frac{\mathbf{1}}{\mathbf{7}}, \frac{\mathbf{1}}{\mathbf{7}}\right)$ has $||\boldsymbol{v}^*||_{\boldsymbol{\infty}} = \frac{\mathbf{1}}{\mathbf{7}}$



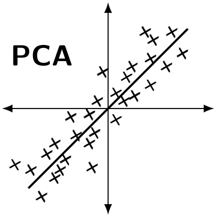
$\boldsymbol{\ell}^\infty$ square

Basis pursuit Minimize $|x_1| + \cdots + |x_n|$ subject to $A\boldsymbol{x} = \boldsymbol{b}$

LASSO with noise Minimize $||A\boldsymbol{x} - \boldsymbol{b}||^2 + \lambda\,\Sigma|x_i|$

LASSO with penalty Minimize $||A\boldsymbol{x} - \boldsymbol{b}||^2$ with $\Sigma|x_i| \le L$

Good ADMM algorithms alternate $\boldsymbol{\ell^2}$ problem and $\boldsymbol{\ell^1}$ problem



A is $\mathbf{2} \times \boldsymbol{n}$ (large nullspace)

AA^{T} is $\mathbf{2} \times \mathbf{2}$ (small matrix)

$A^{\mathrm{T}}A$ is $\boldsymbol{n} \times \boldsymbol{n}$ (large matrix)

Two singular values $\boldsymbol{\sigma_1} > \boldsymbol{\sigma_2} > \mathbf{0}$

The sample covariance matrix is defined by $S = \dfrac{\boldsymbol{AA^{\mathrm{T}}}}{\boldsymbol{n-1}}$**.**

The sum of squared distances from the data points to the u_1 line is a minimum.

Total variance $\qquad \boldsymbol{T = (\sigma_1^2 + \cdots + \sigma_r^2)/(n-1)}$**.**