FIVE FACTORIZATIONS

OF A MATRIX

Gilbert Strang (2023)

Introduction to Linear Algebra (6th ed)

KEY IDEAS AT THE START

Linear independence and dependence

$A\boldsymbol{x} =$ combination of columns of $A = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$

Each column of CR is a combination of the columns of C

$A = CR$: Column space from C, row space from R

1 $\boldsymbol{A} = \boldsymbol{CR}$ $\quad (m \times r)\,(r \times n)$ for any m by n matrix A of rank r

2 $\boldsymbol{A} = \boldsymbol{LU}$ or $\boldsymbol{PA} = \boldsymbol{LU}$ $\qquad L =$ lower triangular

$U =$ upper triangular $\qquad P =$ permutation

For n by n invertible matrices, solve $\boldsymbol{Ax} = \boldsymbol{b}$ in 2 steps

Triangular $\boldsymbol{Lc} = \boldsymbol{b}$ Triangular $\boldsymbol{Ux} = \boldsymbol{c}$

Then $\boldsymbol{Ax} = \boldsymbol{LUx} = \boldsymbol{Lc} = \boldsymbol{b}$

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 3 & 7 \\ 1 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix} = CR$$

Columns 1 and 2 of A are independent: rank 2

Those independent columns go into C

Column 3 of $A = 2$(column 1) $+ 1$(column 2)

So 2 and 1 go into column 3 of $R = \begin{bmatrix} I & F \end{bmatrix}$

The $n - r$ dependent columns of A
are combinations CF of the r independent columns

Every $A = CR = C \begin{bmatrix} I & F \end{bmatrix} P =$
$\Big[$ Independent cols Dependent cols $\Big]$ Permute cols

Suppose the r columns of C are columns $j_1, \ldots j_r$ of A

Then P puts the r columns of I into columns $j_1, \ldots j_r$ of R

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = CR$$

$$R = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} P$$

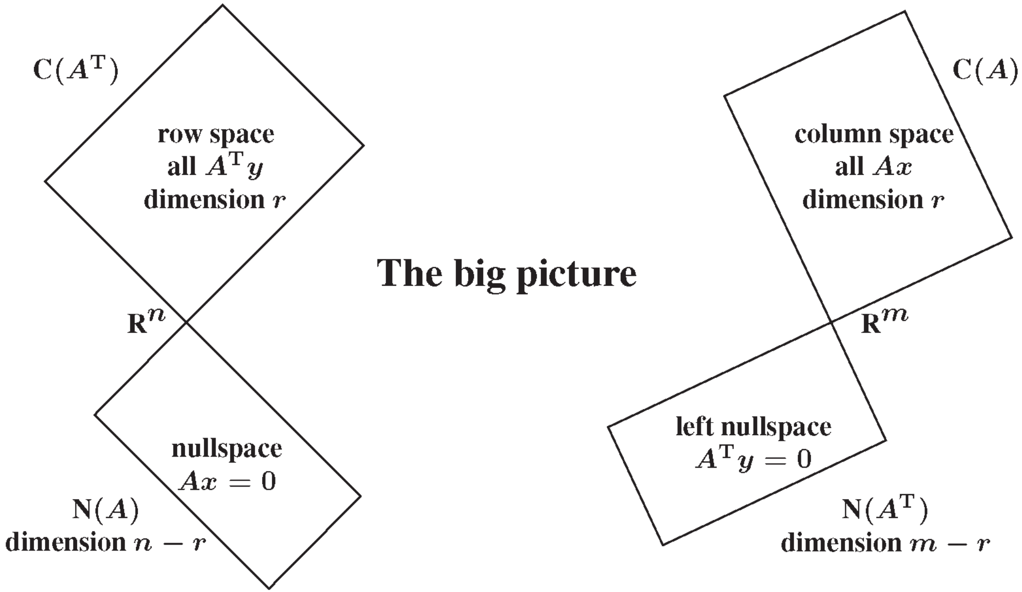
C has r independent columns // R has r independent rows

From $A = CR$, columns of C span the column space of A

From $A = CR$, the rows of R span the row space of A

Column space and row space of A have same dimension r

The Big Picture : Four Subspaces



Elimination computes the reduced row echelon form of A

$$\textbf{rref}(A) = \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} P$$

Finding R first

I locates indep. columns

First $k+1$ columns: $\begin{bmatrix} I_k & F_k \\ 0 & 0 \end{bmatrix} P_k$ are followed by $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}$

If $\boldsymbol{\ell} =$ zero vector then $\boldsymbol{u}$ joins F_k to produce F_{k+1}

If $\boldsymbol{\ell} \neq \mathbf{0}$ then $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}$ becomes $\begin{bmatrix} \mathbf{0} \\ 1 \\ \mathbf{0} \end{bmatrix}$ to produce I_{k+1}

Note: Every part I, F, P of $\textbf{rref}(A)$ is determined by A

$$
\begin{aligned}
x_1 + 2x_2 + 11x_3 + 17x_4 = 0 &\longrightarrow & x_1 + 3x_3 + 5x_4 &= 0 \\
3x_1 + 7x_2 + 37x_3 + 57x_4 = 0 &\longrightarrow & x_2 + 4x_3 + 6x_4 &= 0 \\
4x_1 + 9x_2 + 48x_3 + 74x_4 = 0 &\longrightarrow & 0 &= 0 \\
\mathsf{A}\boldsymbol{x} = \mathbf{0} &\longrightarrow & \mathsf{R}\boldsymbol{x} &= \mathbf{0}
\end{aligned}
$$

$$
\mathsf{A} = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \\ 4 & 9 & 48 & 74 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix} = CR
$$

Nullspace basis $\begin{bmatrix} -3 \\ -4 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -5 \\ -6 \\ 0 \\ 1 \end{bmatrix}$ = Solve $R\boldsymbol{x} = \mathbf{0}$ to find "special solutions to $A\boldsymbol{x} = \mathbf{0}$"

Block elimination $$P_r A P_c = \begin{bmatrix} W & H \\ J & K \end{bmatrix} \to \begin{bmatrix} I & W^{-1}H \\ 0 & 0 \end{bmatrix}$$

The "intersection" of r independent rows of A

with r independent columns of A

produces an r by r invertible matrix W

Elimination reduces W to I and finds $R = \begin{bmatrix} I & W^{-1}H \end{bmatrix}$

Factorization 3 "Gram-Schmidt"

$\boldsymbol{A} = \boldsymbol{QR} =$ (orthogonal) (triangular)

From independent columns in A to
orthogonal unit vectors $\boldsymbol{q}_1$ to $\boldsymbol{q}_n$

$$Q^{\mathrm{T}}Q = \begin{bmatrix} & \boldsymbol{q}_1^{\mathrm{T}} & \\ \cdot & \cdots & \cdot \\ & \boldsymbol{q}_n^{\mathrm{T}} & \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{bmatrix} = I$$

Why is Q so good? Least squares finds $\widehat{\boldsymbol{x}}$ to minimize $||A\boldsymbol{x} - \boldsymbol{b}||^2$

$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \to R^{\mathrm{T}}Q^{\mathrm{T}}QR\widehat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b} \to R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ is easy to solve

Eigenvalues	$S\boldsymbol{x} = \lambda\boldsymbol{x}$	Singular values	$A\boldsymbol{v} = \sigma\boldsymbol{u}$
$S = S^{\mathrm{T}}$	Orthogonal eigenvectors $\boldsymbol{x}$	Orthogonal	$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$
		Orthogonal	$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$

$$S\begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \qquad A\begin{bmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 \cdots \boldsymbol{u}_r \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

$$\boldsymbol{SX = X\Lambda} \quad \boldsymbol{S = X\Lambda X^{-1} = X\Lambda X^{\mathrm{T}}} \quad \boldsymbol{AV = U\Sigma} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}}$$

Every matrix $A =$ (rotation U) (stretching Σ) (rotation V^{T})

($\boldsymbol{v}$'s are eigenvectors of $A^{\mathrm{T}}A$) ($\boldsymbol{u}$'s are eigenvectors of AA^{T})

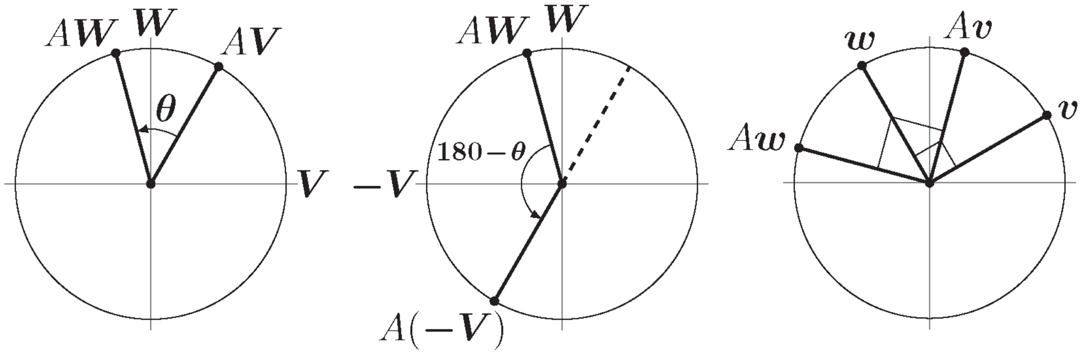


Figure: Angle θ from AV to AW is below 90°. Angle $180-\theta$ from AW to $-AV$ is above 90°. Somewhere in between, as v moves from V toward W, the angle from Av to Aw is exactly 90°. The pictures don't show vector lengths.

Interpolation / Approximation using Deep Learning ?

Given : Values w_k at N points u_k in $\mathbf{R}^n$

Goal : Learning function $F(x, u)$ with weights x and $F(x, u_k) \approx w_k$

Chain of functions
$F(x, u) = F_L(x_L, F_{L-1}(x_{L-1}, F_{L-2}(\ldots F_1(x_1, u))))$

Composition of L simple functions F_k : vector $\longrightarrow$ vector

The weights $\boldsymbol{x} = x_L, x_{L-1}, \ldots, x_1$ are chosen so that
$F(x, u_k) \approx w_k$ Each $x_k =$ matrix A_k and vector b_k

$$F_k(x_k, F_{k-1}) = \text{ReLU}(A_k F_{k-1} + b_k)$$

Nonlinear $\text{ReLU}(y) = \begin{matrix} y & y \geq 0 \\ 0 & y \leq 0 \end{matrix}$ applied to each component

Compute weights x to minimize the loss $||F(x, u) - w||$

Gradient descent or Stochastic gradient descent : not Newton

Resource: A Vision of Linear Algebra
Gilbert Strang