The Column-Row Factorization $A = CR$

A new start for linear algebra

Gilbert Strang

MIT

Linear Algebra for Everyone (2020)

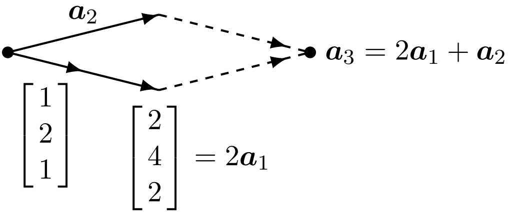
$$A = \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{5} \\ \mathbf{2} & \mathbf{3} & \mathbf{7} \\ \mathbf{1} & \mathbf{3} & \mathbf{5} \end{bmatrix} \qquad \begin{array}{l} m = 3 \text{ rows} \\ n = 3 \text{ columns} \end{array}$$

Are the columns independent ? Go left to right

Column 1 OK   Column 2 OK   Column 3 ?

Column $3 = 2\,$(Column 1) $+1\,$(column 2) **Dependent**

Column 3 is in the plane of Columns 1 and 2



Matrix $C = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix}$ of independent columns in $A = \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{5} \\ \mathbf{2} & \mathbf{3} & \mathbf{7} \\ \mathbf{1} & \mathbf{3} & \mathbf{5} \end{bmatrix}$

The matrix A has **column rank** $\boldsymbol{r = 2}$

The **column space** of A is a plane in $\mathbf{R}^3$

The column space contains all combinations of the columns

Column space of $A =$ Column space of C ((but $A \neq C$))

Express the steps by multiplications Ax and CR

$Ax =$ **matrix times vector** $=$ **combination of columns of A**

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 3 & 7 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 2\,(\text{Column } 1) + 1\,(\text{Column } 2) - 1\,(\text{Column } 3)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \text{(dot products of } \boldsymbol{x} \text{ with rows of } A)$$

$\boldsymbol{CR} =$ **Matrix times matrix** $= \boldsymbol{C}$ **times each column of $\boldsymbol{R}$**

Use dot products (low level) or take combinations of the columns of C

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 3 & 7 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & \mathbf{2} \\ 0 & 1 & \mathbf{1} \end{bmatrix} \quad \text{is } \boldsymbol{A = CR}$$

Check C times each column of R

$$\begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{matrix} 2\,(\text{Column } 1) + (\text{Column } 2) \\ 2\boldsymbol{a}_1 + \boldsymbol{a}_2 = \boldsymbol{a}_3 \end{matrix} = \begin{bmatrix} 5 \\ 7 \\ 5 \end{bmatrix}$$

How to find CR for every A? **Elimination !**

$\boldsymbol{A} = \boldsymbol{CR}$ is $(m$ by $n) = (m$ by $r)\,(r$ by $n)$

$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P}$ and $\boldsymbol{A} = \boldsymbol{CR} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{CF} \end{bmatrix} \boldsymbol{P}$

In reality we compute R before C !! The columns of I in R tell us the independent columns of A in C.

The permutation P puts those columns in the right places (if they are not the first r columns of A)

$\boldsymbol{R} =$ reduced row echelon form rref(A) (zero rows removed)

Here are the steps to establish $A = CR$

We know $EA = \textbf{rref}(A)$ and $A = E^{-1}\,\textbf{rref}(A)\ :\ E$ is $m \times m$

Remove $m - r$ zero rows from $\textbf{rref}(A)$ and $m - r$ columns from E^{-1}

This leaves $A = C\begin{bmatrix} I & F \end{bmatrix} P = CR$ Dependent columns of A are $\boldsymbol{CF}$

C has r independent columns $\qquad R$ has r independent rows

Rows of $A = CR$ are combinations of the rows of R

Row space of A = Row space of R !

If A has 2 independent columns in C then **$\boldsymbol{A}$ has 2 independent rows in $\boldsymbol{R}$**

Column rank = Row rank = $\boldsymbol{r}$ GREAT THEOREM

Look at $A = CR$ both ways: Combine columns of C Combine rows of R

$\boldsymbol{r = 1}$ Rank one matrix $A = (1 \text{ column})\,(1 \text{ row})$

$$\begin{bmatrix} 1 & 2 & 10 & 100 \\ 2 & 4 & 20 & 200 \\ 1 & 2 & 10 & 100 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 10 & 100 \end{bmatrix} = CR$$

If the column space is a line in 3-dimensional space

then the row space is a line in 4-dimensional space

A adds up (Column k of C) (Row k of R) $=$ **New way to multiply $\boldsymbol{CR}$**

Rank r matrix $=$ Sum of r matrices of rank 1

Geometry of A: Four Fundamental Subspaces

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

Row space $\mathbf{C}(A^{\mathrm{T}})$ = all combinations of columns of A^{T} = all $A^{\mathrm{T}}\boldsymbol{y}$

Nullspace $\mathbf{N}(A)$ = all solutions $\boldsymbol{x}$ to $A\boldsymbol{x} = \mathbf{0}$

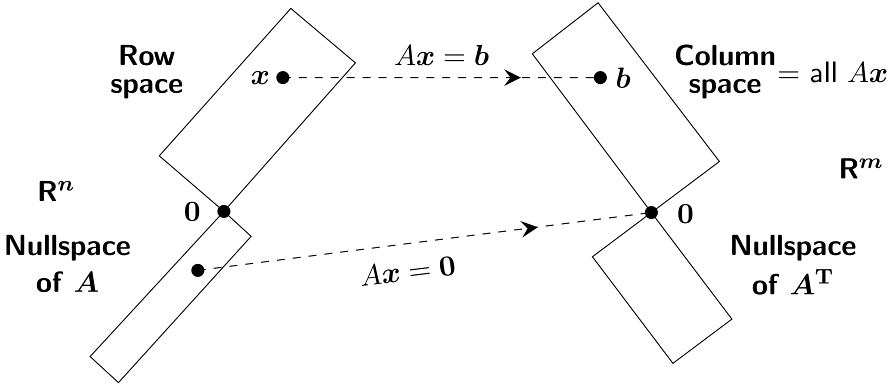
Nullspace of A^{T} $\mathbf{N}(A^{\mathrm{T}})$ = all solutions $\boldsymbol{y}$ to $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$

Dimensions $\boldsymbol{r}$ $\boldsymbol{r}$ $\boldsymbol{n-r}$ $\boldsymbol{m-r}$

Row space is orthogonal to nullspace!

$$\begin{bmatrix} \text{row } 1 \\ \cdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \cdot \\ \mathbf{0} \end{bmatrix}$$

m rows and n columns $\qquad$ r independent rows and columns



BIG PICTURE OF LINEAR ALGEBRA

Square invertible matrices $m = n = r$

Nullspaces = zero vector only

Magic factorization $\quad A = CW^{-1}R_*$

$\boldsymbol{C} = r$ independent columns of A $\qquad$ $R_* = r$ independent rows of A

$\boldsymbol{W} = r \times r$ matrix $=$ **intersection of columns in C and rows in $\boldsymbol{R_*}$**

The factorization is just block elimination on A. The block pivot is W.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 3 & 7 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 3 & 7 \end{bmatrix}$$

W is invertible and $\boldsymbol{WR} = \boldsymbol{R_*}$ from r rows of $CR = A$

Randomized linear algebra $\boldsymbol{A \approx CW^{-1}R_*}$

Large matrices / thin samples "Skeleton factors"

References to CUR_*^3 R. Penrose (1956) *On best approximate solutions of linear matrix equations,* Math. Proc. Cambridge Phil. Soc. $\mathbf{52}$ 1719-.

Hamm and Huang (2020) *Perspectives on* CUR *Decompositions* arXiv 1907.12668 and ACHA $\mathbf{48}$

Goreinov, Tyrtyshnikov, and Zamarashkin (1997) *Pseudoskeleton approximation* LAA 261

Martinsson and Tropp (2020) *Randomized numerical linear algebra : Foundations and Algorithms* Acta Numerica and arXiv : 2002.01387

Randomized Numerical Linear Algebra $\boldsymbol{A \approx CUR}$

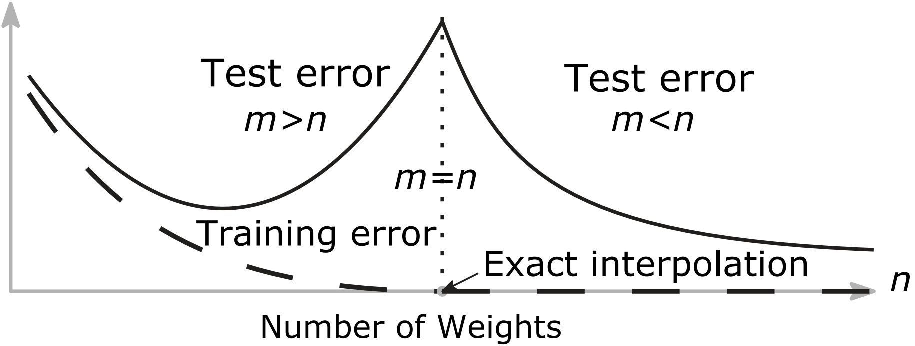
Famous Factorizations of a Matrix

$$\begin{aligned}
\boldsymbol{A} &= \boldsymbol{LU} && = \text{(lower triangular } L\text{)}\,\text{(upper triangular } R\text{)} \\
\boldsymbol{A} &= \boldsymbol{QR} && = \text{(orthogonal columns in } Q\text{)}\,\text{(upper triangular } R\text{)} \\
\boldsymbol{S} &= \boldsymbol{Q\Lambda Q^{\mathrm{T}}} && = \text{(eigenvectors in } Q\text{)}\,\text{(eigenvalues in } \Lambda\text{)} \\
\boldsymbol{A} &= \boldsymbol{U\Sigma V^{\mathrm{T}}} && = \text{(singular vectors in } U \text{ and } V\text{)}\,\text{(singular values in } \Sigma\text{)} \\
A\boldsymbol{v}_k &= \sigma_k \boldsymbol{u}_k && \ \text{(orthogonal vectors } \boldsymbol{v} \text{ mapped to orthogonal vectors } \boldsymbol{u}\text{)}
\end{aligned}$$

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix} \qquad \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Full rank $r = m = n$	$r = n$ indep. columns	$r = m$ indep. rows
$\boldsymbol{A}$ is invertible	$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible	$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is invertible
$\begin{bmatrix} & A & \\ & & \end{bmatrix}$	$\begin{bmatrix} A \\ \\ \\ \end{bmatrix}$	$\begin{bmatrix} & & A & & \\ & & & & \end{bmatrix}$
Solve $A\boldsymbol{x} = \boldsymbol{b}$	$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$	$AA^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{b} \to \overline{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{y}$
$\boldsymbol{x}$ **exact** solution	$\widehat{\boldsymbol{x}}$ **least squares** solution	$\overline{\boldsymbol{x}}$ **minimum norm** solution

The minimum norm solution $\overline{\boldsymbol{x}}$ has no nullspace component / use the pseudoinverse $\overline{\boldsymbol{x}} = A^{+}\boldsymbol{b}$



Deep learning has found that overfitting can help ! A big question in the theory of neural networks using ReLU

Video Lectures ocw.mit.edu/courses/mathematics YouTube/mitocw

Math 18.06 Linear Algebra (including 2020 Vision)

Math 18.065 Deep Learning

Books

Introduction to Linear Algebra, (2016)	math.mit.edu/linearalgebra
Linear Algebra & Learning from Data (2019)	math.mit.edu/learningfromdata
Linear Algebra for Everyone (2020)	math.mit.edu/everyone