LU Factorization with Pivoting

Not every $A \in \mathbb{R}^{n\times n}$ has an LU factor.

e.g. $\begin{bmatrix} 0 & -1 \\ 4 & 5 \end{bmatrix}$

$$A = A^{(1)} \rightsquigarrow A^{(2)} \rightsquigarrow A^{(3)} \cdots \rightsquigarrow A^{(n)} = U$$

$A^{(j)}$ – 0s are introduced on columns $1, \ldots, j-1$

$a_{jj}^{(j)} \approx 0 \implies |M_{kj}|$ are large $(k>j)$

pivot $\implies$ Severe rounding errors

(Partial) Pivoting Strategy

$$A^{(j)} \rightsquigarrow A^{(j+1)}$$

(1) Find k s.t.

$$|a_{kj}^{(j)}| = \max_{j \le i \le n} |a_{ij}^{(j)}|$$

(2) Swap rows k and j.

do these before introducing 0s on the jth column

$a_{jj}^{(j)}$

$\otimes$

x

o

o

o

x

↑ jth column

Ex

$$\begin{bmatrix} 2 & 0 & 4 \\ 1 & 2 & 1 \\ -3 & 1 & 5 \end{bmatrix} \underset{1 \leftrightarrow 3}{\sim} \begin{bmatrix} -3 & 1 & 5 \\ 1 & 2 & 1 \\ 2 & 0 & 4 \end{bmatrix}$$

$$\underset{\substack{\mu_{21} = 1/3 \\ \mu_{31} = 2/3}}{\sim} \begin{bmatrix} -3 & 1 & 5 \\ 0 & 7/3 & 8/3 \\ 0 & 2/3 & 22/3 \end{bmatrix}$$

$$\underset{\mu_{32} = -2/7}{\sim} \begin{bmatrix} -3 & 1 & 5 \\ 0 & 7/3 & 8/3 \\ 0 & 0 & 138/21 \end{bmatrix}$$

Row Interchange

$P^{(ij)}$ - obtained from I by swapping rows i & j

e.g. $P^{(31)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

$P^{(31)}A$ - swaps rows 1 & 3

$P^{(ij)}A$ - swaps rows i & j of A.

Permutation Matrices

Matrices with only one nonzero entry that is equal to one along every row and every column.

* Row interchange matrix is a special permutation matrix
* Products of row interchange matrices are permutation matrices

$$\text{e.g.}\quad \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{\substack{\text{Permutation}\\ \text{matrix}}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P^{(21)}} \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{P^{(31)}}$$

Proposition

A permutation matrix $P \in \mathbb{R}^{n \times n}$ is orthogonal, i.e., $P^T P = I_n$.

Proof

$$[P^T P]_{ij} = \begin{cases} p_i^T p_j = 0 & \text{if } i \neq j \\ p_j^T p_j = 1 \end{cases} \qquad \square$$

THM

Every matrix $A \in \mathbb{R}^{n \times n}$ has a factorization of the form

$$PA = LU$$

where $P \in \mathbb{R}^{n \times n}$ is a permutation matrix,
$L \in \mathbb{R}^{n \times n}$ is unit lower triangular,
$U \in \mathbb{R}^{n \times n}$ is upper triangular.

Proof

By induction on n.

Base case, $n=2$.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

~~if~~ suppose we apply partial pivoting strategy. If $|a| \geq |c|$ (and assume $a \neq 0$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & d - bc/a \end{bmatrix}$$

Otherwise (again assuming $c \neq 0$)

$$\underbrace{\begin{bmatrix} c & d \\ a & b \end{bmatrix}}_{P^{(21)} A} = \begin{bmatrix} 1 & 0 \\ a/c & 1 \end{bmatrix} \begin{bmatrix} c & d \\ 0 & b - ad/c \end{bmatrix}.$$

Finally, if $a = c = 0$

$$\begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix}.$$

Inductive case,
Assume every $\hat{A} \in \mathbb{R}^{n \times n}$ has the desired factorization

$$\hat{P}\hat{A} = \hat{L}\hat{U}.$$

For a given $A \in \mathbb{R}^{(n+1)\times(n+1)}$, suppose we apply the partial pivoting strategy which swaps rows 1 & r. We claim

$$(+) \quad P^{(r1)} A = \begin{bmatrix} a_{r1} & \beta^T \\ \alpha & B \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 \\ m & I \end{bmatrix} \begin{bmatrix} u & v^T \\ 0 & C \end{bmatrix} \cdot \quad \begin{array}{l} \exists m, v \in \mathbb{R}^n \\ \exists u \in \mathbb{R} \\ \exists C \in \mathbb{R}^{n\times n} \end{array}$$

If $a_{r1} = 0$, then $\alpha = 0$. In this case, choose $m = 0$, $u = 0$, $v = \beta$ and $C = B$ for the satisfaction of (+).

Otherwise, (+) holds if and only if

$$u = a_{r1}$$
$$v = \beta$$
$$um = \alpha$$
$$mv^T + C = B.$$

Hence, for choices $u = a_{r1} \neq 0$, $v = \beta$, $m = \alpha / u$ and $C = B - mv^T$ (+) holds.

By the inductive hypothesis

$$\hat{P} C = \hat{L} \hat{U}$$

for some $\hat{P} \in \mathbb{R}^{n\times n}$ that is a permutation matrix,
$\hat{L} \in \mathbb{R}^{n\times n}$ unit lower triangular,
$\hat{U} \in \mathbb{R}^{n\times n}$ upper triangular.

Plugging this in (+)

$$P^{(r-1)}A = \begin{bmatrix} I & 0 \\ m & I \end{bmatrix} \begin{bmatrix} u & v^T \\ 0 & \hat{P}^T \hat{L} \hat{U} \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 \\ m & \hat{P}^T \end{bmatrix} \begin{bmatrix} u & v^T \\ 0 & \hat{L} \hat{U} \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 \\ 0 & \hat{P}^T \end{bmatrix} \begin{bmatrix} I & 0 \\ \hat{P}m & I \end{bmatrix} \begin{bmatrix} u & v^T \\ 0 & \hat{L} \hat{U} \end{bmatrix}$$

$$\Longrightarrow \underbrace{\begin{bmatrix} I & 0 \\ 0 & \hat{P} \end{bmatrix} P^{(r-1)}}_{P} A = \underbrace{\begin{bmatrix} I & 0 \\ \hat{P}m & \hat{L} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} u & v^T \\ 0 & \hat{U} \end{bmatrix}}_{U}$$

as desired. $\square$

Remark

The proof suggests that we can apply the usual algorithm with row swaps. But

(1) keep the permutation matrix (row-interchange info) at every iteration

(2) $\hat{P}m$ at bottom left of L suggests the swap of the multiplier at the previous columns that are already computed.

Ex (see example on page 2)

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 4 \\ 1 & 2 & 1 \\ -3 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -2/3 & +2/7 & 1 \end{bmatrix} \begin{bmatrix} -3 & 1 & 5 \\ 0 & 7/3 & 8/3 \\ 0 & 0 & 138/21 \end{bmatrix}$$

Solution of Linear Systems
Using LU factorization

$$Ax = b$$
$$PAx = Pb$$
$$L\underbrace{Ux}_{y} = Pb$$

(1) Compute $PA = LU$
$\sim 2n^3/3$ operations

(2) Solve $Ly = Pb$
by forward substitution
$\sim n^2$ operations

(3) Solve $Ux = y$
by back substitution
$\sim n^2$ operations

Especially useful if

$$Ax = b_1, \ \ldots \ Ax = b_k$$

need to be solved for several $b_1, \ldots, b_k \in \mathbb{R}^n$.

<u>Ex</u> (Inverse Iteration)

Consider the sequence $\{\underline{y}_k\}$

$$\underline{y}_{k+1} = A^{-1}\underline{y}_k$$

for an invertible $A \in \mathbb{R}^{n \times n}$ and given $\underline{y}_0 \in \mathbb{R}^n$.

Compute $PA = LU$ only once initially.

Every iteration is $\sim 2n^2$ operations.

For a symmetric A, generically

$$\lim_{k \to \infty} \underline{y}_k^T A \underline{y}_k / \|\underline{y}_k\|^2 = \underbrace{\lambda_{min}(A)}_{\substack{\text{smallest} \\ \text{eigenvalue} \\ \text{of } A \text{ in } |\cdot|}}.$$

Norms and Condition Numbers

$$(*) \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 10^{-8} \end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{b}, \qquad x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 10^{-8} \end{bmatrix}}_{A} \hat{x} = \underbrace{\begin{bmatrix} 1 \\ 10^{-8} \end{bmatrix}}_{\hat{b}}, \qquad \hat{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$b \approx \hat{b}$, but $x \not\approx \hat{x}$

Linear system and A are said to be ill-conditioned. ("ill-conditioning" will be defined formally soon.)

Normed vector spaces

$\mathbb{R}$ with $|\cdot|$

$|\cdot|$ satisfies (for all $v, w \in \mathbb{R}$)

(i) $|\alpha v| = |\alpha|\,|v|$ for all $\alpha \in \mathbb{R}$

(ii) $|v| > 0$ unless $v = 0$

(ii) $|v + w| \leq |v| + |w|$

Defn

Let V be a vector space over the field of $\mathbb{R}$. A norm $\|\cdot\|: V \to \mathbb{R}$ satisfies

(POSITIVITY) (i) $\|v\| > 0$ for all $v \in V$, $v \neq 0$,

(HOMOGENEITY) (ii) $\|\alpha v\| = |\alpha| \, \|v\|$ for all $v \in V$, all $\alpha \in \mathbb{R}$

(TRIANGULAR INEQUALITY) (iii) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$

V together with $\|\cdot\|$ is called a normed vector space.

Common norms in $\mathbb{R}^n$

Let $v \in \mathbb{R}^n$

2-norm $\|v\|_2 := \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$

1-norm $\|v\|_1 := |v_1| + |v_2| + \dots + |v_n|$

∞-norm $\|v\|_\infty := \max_{j=1,\dots,n} |v_j|$

(Note: $\|v\|_1 \geq \|v\|_2 \geq \|v\|_\infty$)

Ex

$$v = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}$$

$$\|v\|_2 = \sqrt{1^2 + (-4)^2 + (2)^2} = \sqrt{21}$$

$$\|v\|_1 = |1| + |-4| + |2| = 7$$

$$\|v\|_\infty = \max\{|1|, |-4|, |2|\} = 4$$

The 2-norm is indeed a norm on $\mathbb{R}^n$.

Positivity

$v \in \mathbb{R}^n, \; v \neq 0$

say $v_j \neq 0$

$$\|v\|_2 = \sqrt{v_1^2 + \dots + v_n^2} \geq |v_j| > 0.$$

Homogeneity

$v \in \mathbb{R}^n, \; \alpha \in \mathbb{R}$

$$\|\alpha v\|_2 = \sqrt{(\alpha v_1)^2 + \dots + (\alpha v_n)^2}$$

$$= \sqrt{\alpha^2 (v_1^2 + \dots + v_n^2)}$$

$$= |\alpha| \sqrt{v_1^2 + \dots + v_n^2} = |\alpha| \, \|v\|_2$$

Triangle inequality

Follows from the Cauchy-Schwarz inequality

THM

Let $v, w \in \mathbb{R}^n$. Then

$$\left| \sum_{j=1}^{n} v_j w_j \right| \leq \|v\|_2 \, \|w\|_2$$

Proof

Consider

$$p(t) = \|v + tw\|_2^2$$

$$= \sum_{j=1}^{n} (v_j + t w_j)^2$$

A nonnegative quadratic polynomial in t with at most one real root.

$$p(t) = \sum_{j=1}^{n} v_j^2 + 2t \sum_{j=1}^{n} v_j w_j + t^2 \sum_{j=1}^{n} w_j^2$$

must have its disriminant nonpositive (otherwise $p(t)$ would have two real roots), i.e.,

$$4t^2 \left(\sum_{j=1}^{n} v_j w_j \right)^2 - 4t^2 \sum_{j=1}^{n} v_j^2 \sum_{j=1}^{n} w_j^2 \leqslant 0$$

$$\Longrightarrow \quad \left| \sum_{j=1}^{n} v_j w_j \right| \leqslant \sqrt{\sum_{j=1}^{n} v_j^2 \sum_{j=1}^{n} w_j^2}$$

$$= \|v\|_2 \, \|w\|_2 . \qquad \square$$

It follows that

$$\|v + w\|_2^2 = \sum_{j=1}^{n} (v_j + w_j)^2$$

$$= \sum_{j=1}^{n} v_j^2 + 2 \sum_{j=1}^{n} v_j w_j + \sum_{j=1}^{n} w_j^2$$

$$\overset{\text{by Cauchy-Schwarz inequality}}{\leq} \|v\|_2^2 + 2 \|v\|_2 \|w\|_2 + \|w\|_2^2$$

$$= \left(\|v\|_2 + \|w\|_2 \right)^2$$

$$\Longrightarrow \quad \|v + w\|_2 \leqslant \|v\|_2 + \|w\|_2$$

p-norm on $\mathbb{R}^n$ $(p \geq 1)$

$$\|v\|_p := \sqrt[p]{|v_1|^p + \dots + |v_n|^p}$$

Ex $v = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}$, $\|v\|_3 = \sqrt[3]{1^3 + 4^3 + 2^3}$
$= \sqrt[3]{25}$

1-norm and 2-norm
are special cases of p-norm

∞-norm is the limit as $p \to \infty$, i.e.,

$$\|v\|_\infty = \lim_{p \to \infty} \|v\|_p$$

in particular define $\tilde{v} := v / \|v\|_\infty$, then

$$1 \leqslant \|\tilde{v}\|_p \leqslant n^{1/p}$$

$$\implies \lim_{p \to \infty} \|\tilde{v}\|_p = 1$$

$$\implies \lim_{p \to \infty} \|v\|_p = \|v\|_\infty .$$

The p-norm for every $p \geqslant 1$ is indeed a norm on $\mathbb{R}^n$.

The triangle inequality follows from the following generalization of the Cauchy-Schwarz inequality.

THM (Hölder's inequality)

Let $p, q \in \mathbb{R}$ s.t. $p, q \geq 1$ and $1/p + 1/q = 1$.
For every $v, w \in \mathbb{R}^n$, we have

$$\left| \sum_{j=1}^{n} v_j w_j \right| \leq \|v\|_p \|w\|_q.$$