MA214:

Mathematics and Contemplations On Linear Algebra and Its Applications

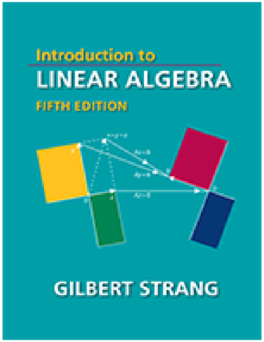
towards building a “Data Scientist”

Waleed A. Yousef, Ph.D.,

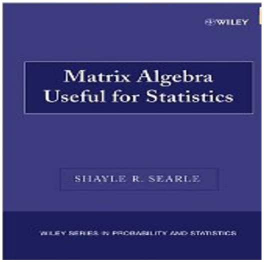
Human Computer Interaction Lab.,
Computer Science Department,
Faculty of Computers and Information,
Helwan University,
Egypt.

March 24, 2019

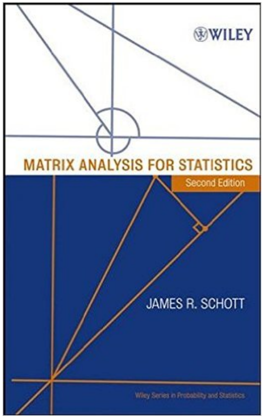
Lectures Notes (http://www.helwan.edu.eg/university/staff/Dr.WaleedYousef/HTML/Home.html) follows:



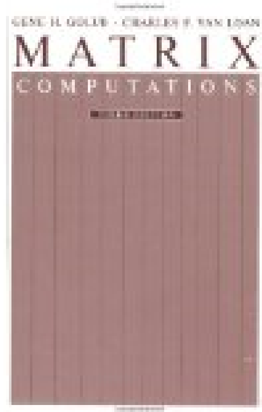
Strang, G., 2016. Introduction to linear algebra, 5th Edition. Wellesley-Cambridge Press



Searle, S. R., 1982. Matrix algebra useful for statistics. Wiley, New York



Schott, J. R., 2005. Matrix analysis for statistics, 2nd Edition. Wiley, Hoboken, N.J



Golub, G. H., Van Loan, C. F., 1996. Matrix computations, 3rd Edition. Johns Hopkins University Press, Baltimore

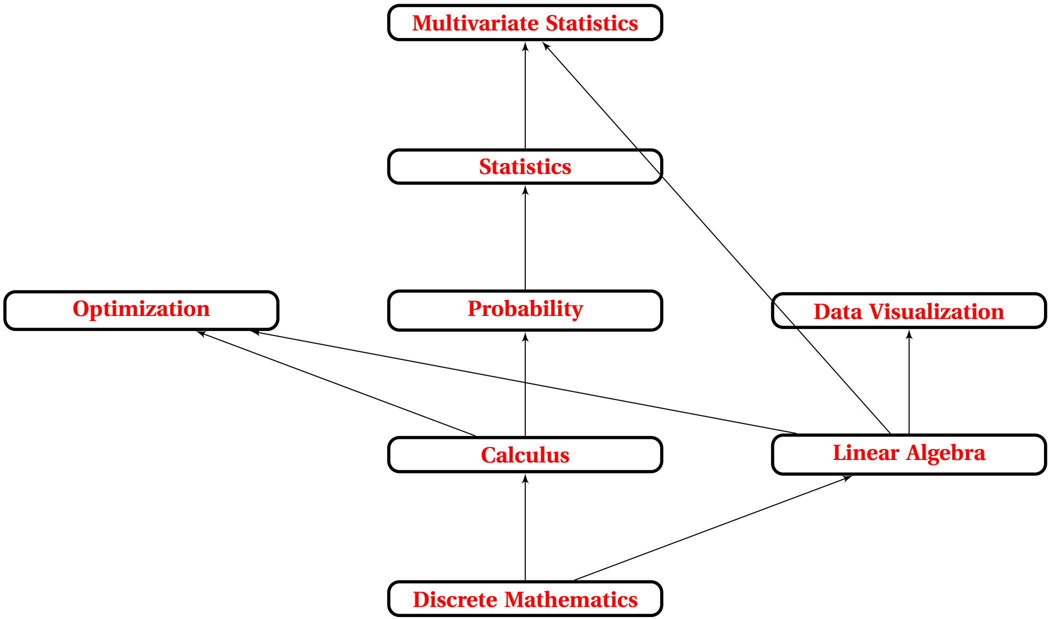
Linear Algebra: FCIHOCW vs. MITOCW

- Arabic vs. English.
- More rigorous treatment.
- Teaching, while “Data Science” in mind.

Course Objectives

- Developing rigorous treatment.
- Developing mathematical foundations to many courses and areas, in particular “Data Science”
- Building intuition.
- Linking to CS applications (e.g., Pattern Recognition, Image Processing, etc.)

Linear Algebra, Prerequisites, and Applications



- Some prerequisites are not so strict; others are possible, e.g., GPU, Algorithms, etc.
- It differs from researchers to practitioners; See pattern recognition course and big picture talk.

Computer Graphics

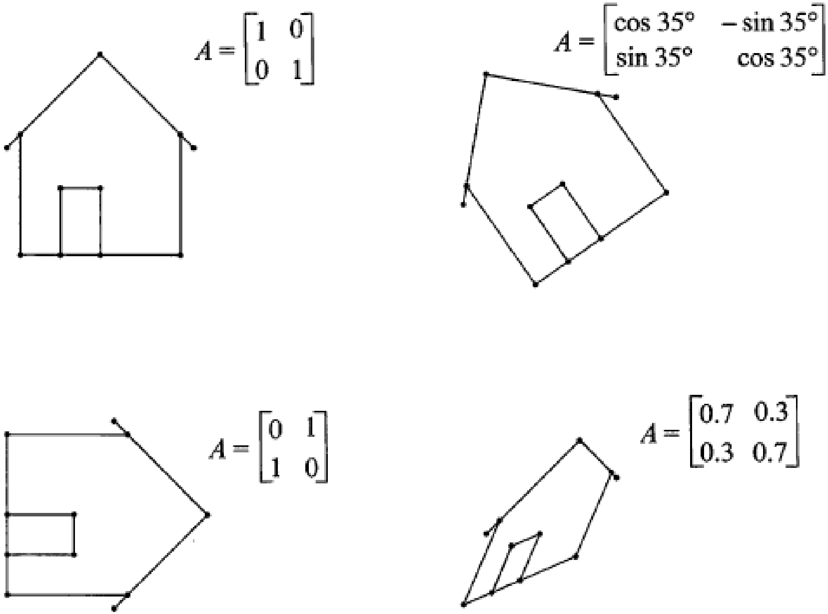


Figure 7.2: Linear transformations of a house drawn by **plot2d**$(A * H)$.

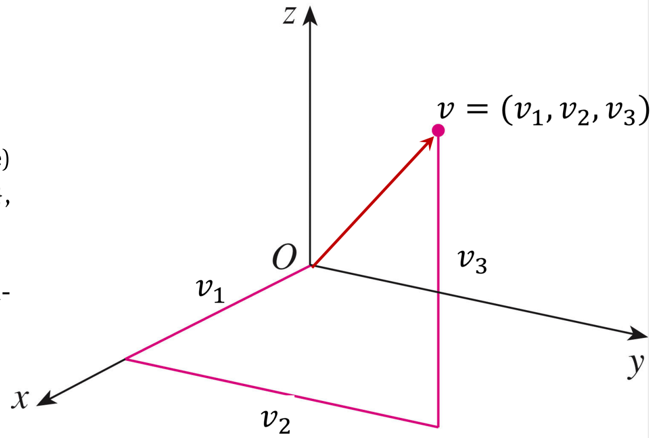
Contents

Chapter 1

Introduction

1.0 Back to School: visual space!

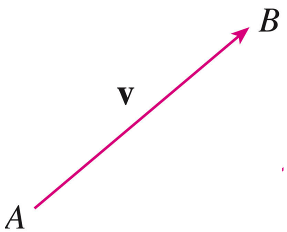
- We locate a point in a 3D space by three numbers.
- The coordinates are perpendicular.
- The order of the axes X, Y, Z: "right-hand" rule.
- The 3-tuple (3 ordered elements, or triple) $(v_1, v_2, v_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 = \{(x, y, z) | x, y, z \in \mathbb{R}\}$, the set of all points.
- The following are equivalent (some books differentiate; we do not):
  - the 3-tuple $v = (v_1, v_2, v_3)$.
  - the point $v = (v_1, v_2, v_3)$.
  - the arrow connecting O to v, i.e., the vector $v = \overrightarrow{Ov} = (v_1, v_2, v_3)$.
- The line segment $\overline{Ov}$ consists of **all** points, not only v.



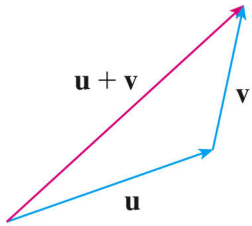
Definition 1 (Geometric Manipulation) .

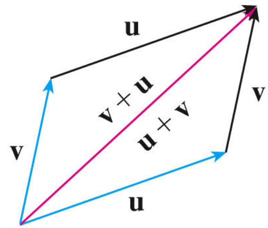
- *A vector is used to indicate a* ***displacement*** *in some* ***direction****;* ***starting point is not important***

- *Start at any point A, move a distance in the direction of* $\overrightarrow{Ov}$*, and end at B. Then,* $\overrightarrow{AB} = \overrightarrow{Ov} = v$*.* ($B \neq \overrightarrow{AB}$*; but* $v = \overrightarrow{Ov}$)

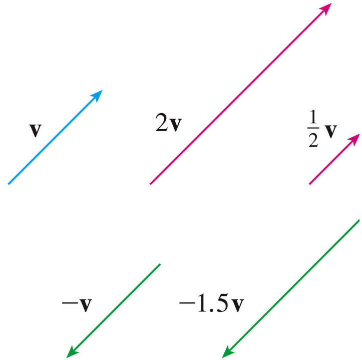


- *Addition:* $u + v$

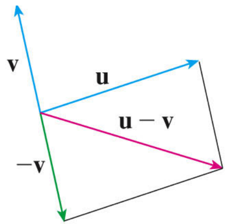


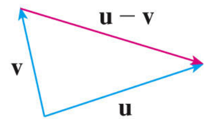


- *Scalar Multiplication: If c is a scalar, then* $u = cv$ *is a vector whose length is* $|c|\times$ *length of v and direction:*



- *Scalar and Addition:*





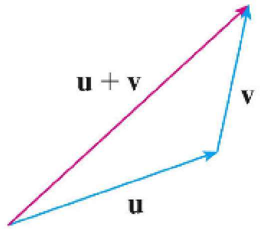
Definition 2 (Algebraic Treatment) . *Addition and Scalar: if* $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3)$:

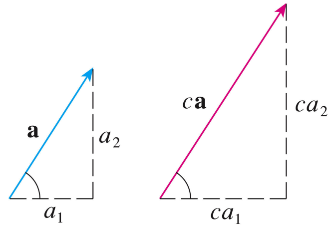
$$\begin{aligned} a + b &= (a_1 + b_1, a_2 + b_2, a_3 + b_3), \\ ca &= (ca_1, ca_2, ca_3). \end{aligned}$$

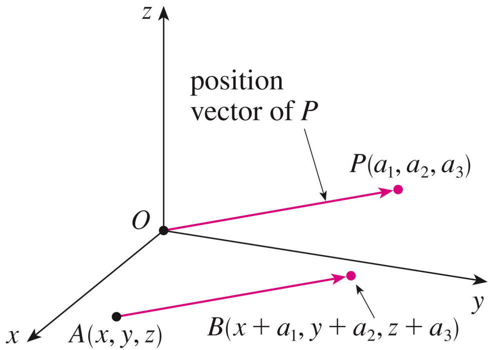
Proof of equivalence. Trivial. ∎

Hint: The displacement is added algebraically: given $P = (a_1, a_2, a_3)$, and any $A = (x, y, z)$. Then:

$$\begin{aligned} P &= \overrightarrow{OP} = \overrightarrow{AB}, \\ B &= A + P = (x + a_1, y + a_2, z + a_3). \end{aligned}$$







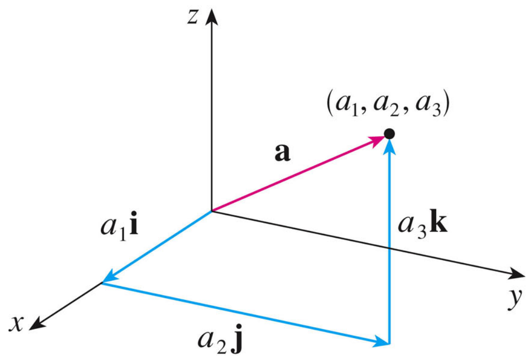
Lemma 3 (Properties of Vectors) . *For any two vectors a and b,*

$$
\begin{aligned}
a+b &= b+a \\
a+(b+c) &= (a+b)+c \\
a+\mathbf{0} &= a \\
a+(-a) &= \mathbf{0} \\
c(a+b) &= ca+cb \\
(c+d)a &= ca+da
\end{aligned}
$$

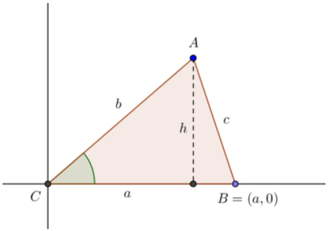
Proof. It is quite straight forward to prove **(HW)** ■

Example 4 *Consider the vector a,*

$$
\begin{aligned}
a &= (a_1, a_2, a_3) \\
&= a_1 i + a_2 j + a_3 k, \\
i &= (1,0,0), \\
j &= (0,1,0), \\
k &= (0,0,1).
\end{aligned}
$$



1.1 Angle, Lengths, and Dot Products (visual space and school again)



- Notation: the vector u, with 3-tuple (u_1, u_2, u_3) is written as: $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ or $u' = (u_1, u_2, u_3)$.

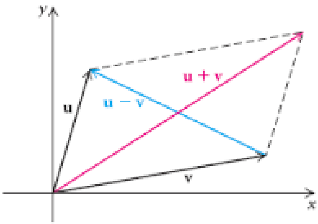
- it is a school business to prove that (whether 2D or 3D):

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\|\cos\theta$$

$$2\|u\|\|v\|\cos\theta = (u_1^2 + u_2^2) + (v_1^2 + v_2^2) - (u_1 - v_1)^2 - (u_2 - v_2)^2 = 2u_1v_1 + 2u_2v_2$$

$$\cos\theta = \frac{u_1v_1 + u_2v_2}{\|u\|\|v\|}.$$



- This is why we defined the dot product to be: $u'v = u_1v_1 + u_2v_2 = \|u\|\|v\|\cos\theta$.

- When $u'v$ is zero we say they are orthogonal.

- If $u = v$, then $\theta = 0$, $u'u = u_1u_1 + u_2u_2 = \|u\|^2$.

- u is **unit vector** if $\|u\| = 1$. Then $\forall u$, $u/\|u\|$ is a unit vector.

$$\begin{aligned} u'v = & \quad \|v\|\|u\|\cos\theta = \|v\| \times \texttt{Projection Length of } u \texttt{ on } v \\ u'(v/\|v\|) = & \quad \texttt{Projection Length of } u \texttt{ on } v \\ v'(u/\|u\|) = & \quad \texttt{Projection Length of } v \texttt{ on } u. \end{aligned}$$

Lemma 5 (Properties) .

- ***Basic properties:***

$$u'v = v'u$$
$$\|au\| = |a|\|u\|$$
$$a(u'v) = (au)'v = au'v$$
$$(au + bv)'w = au'w + bv'w$$
$$(u+v)'(u+v) = u'u + 2u'v + v'v.$$

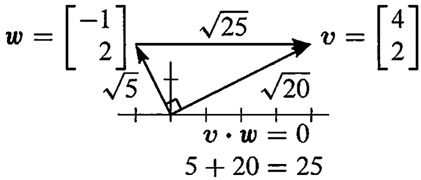
- ***Cauchy-Shwartz inequality:*** $-\|u\|\|v\| \le u'v \le \|u\|\|v\|$
  Proof. *immediate from both:* $-1 \le \cos\theta \le 1$ *and* $u'v = \|u\|\|v\|\cos\theta$. ■

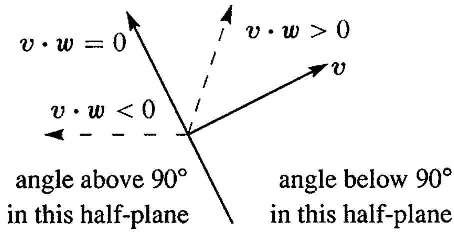
- ***Traingular inequality:*** $\|u+v\| \le \|u\| + \|v\|$
  Proof. $\|u+v\|^2 = (u+v)'(u+v) = u'u + 2u'v + v'v \le \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$. ■

- *Then, we can generalize this definition in higher dimensions, and define the angle between two vectors for* $p > 3$.

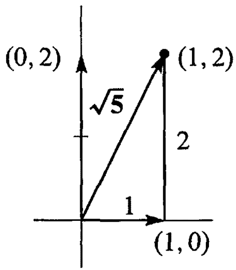
Example 6 $w = (-1, 2)'$, $v = (4, 2)'$, *then*

$$\cos\theta = \frac{w'v}{\|w\|\|v\|} = \frac{(-1)(4) + (2)(2)}{\sqrt{(-1)^2 + (2)^2}\sqrt{(4)^2 + (2)^2}} = \frac{0}{\sqrt{5}\sqrt{20}} = 0$$

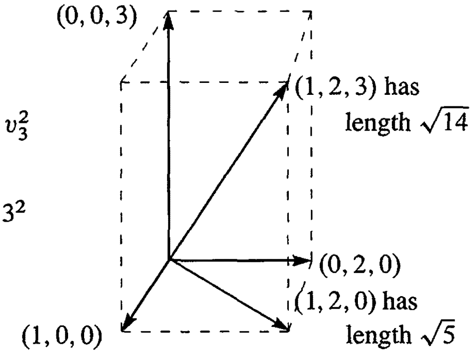




Example 7 (3D) .

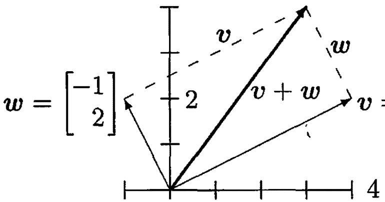


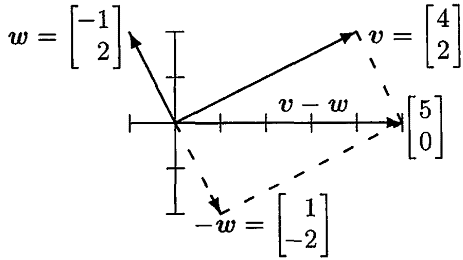
$$\begin{aligned} v \cdot v &= v_1^2 + v_2^2 + v_3^2 \\ 5 &= 1^2 + 2^2 \\ 14 &= 1^2 + 2^2 + 3^2 \end{aligned}$$



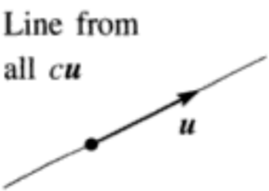
Hint: *To save space, we write, e.g.,* $v = (4, 2)'$. *Sometimes, we drop the prime if there is no confusion.*

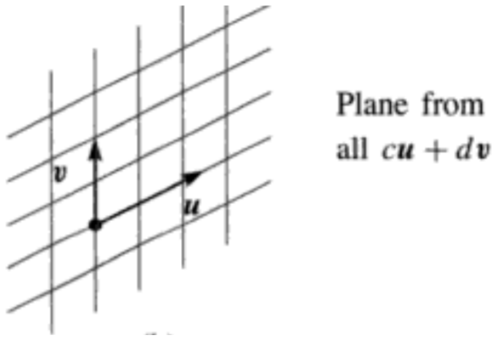
Example 8 (Linear Combination) .





$$v + w = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad v - w = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$





- We can generalize to p-dimensions, although we cannot visualize.
- What is the picture for **ALL** linear combinations? "spanning" the space, independence ...

1.2 Extension and Abstraction: Vectors and Linear Combinations

Extension in both: meaning and number of components to treat applications.

Definition 9 (Vector) *The ordered p-tuple $(v_1, v_2, \cdots, v_p)$, $v_i \in \mathcal{R}$, is called a p-dimensional vector.*

Definition 10 (dot product (inner product), length, angle)

$$\begin{aligned}
\langle u, v\rangle = u.v = u'v &= \begin{pmatrix} u_1, & \cdots, & u_p \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} \\
&= u_1 v_1 + \cdots + u_p v_p = \sum_{i=1}^{p} u_i v_i \\
\|u\| &= \sqrt{u'u} = \sqrt{u_1^2 + u_2^2 + \cdots + u_p^2} \\
\cos\theta &= \frac{u'v}{\|u\|\|v\|}.
\end{aligned}$$

Now, we have to reprove Cauchy-Schwartz inequality; then triangular inequality follows directly!
Proof. $\forall \lambda \in \mathcal{R}$, (we will put it later as $\frac{u'v}{\|v\|^2}$)

$$0 \le \|u - \lambda v\|^2 = \|u\|^2 - 2\lambda u'v + \|\lambda v\|^2 = \|u\|^2 - 2\frac{(u'v)^2}{\|v\|^2} + \frac{(u'v)^2}{\|v\|^4}\|v\|^2 = \|u\|^2 - \frac{(u'v)^2}{\|v\|^2}$$

$$(u'v)^2 \le \|u\|^2\|v\|^2 \quad \Longrightarrow \quad -\|u\|\|v\| \le u'v \le \|u\|\|v\|$$

∎

Definition 11 (Linear Combination: generalization to adding vectors; this is the abstraction) .
Consider the two p-dimensional vectors v and w, and $c,\ d \in R$. We call $cv + dw$ a linear combination.

$$c\begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} + d\begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix} = \begin{pmatrix} cv_1 + dw_1 \\ \vdots \\ cv_p + dw_p \end{pmatrix}.$$

Chapter 2

Solving Linear Equations

2.1 Vectors and Linear Equations

$$\begin{matrix} x-2y=1 \\ 3x+2y=11 \end{matrix} \quad \equiv \quad \begin{pmatrix} 1 & -2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 11 \end{pmatrix} \quad \equiv \quad A\mathbf{x} = \mathbf{b}$$

Column picture (linear combination)

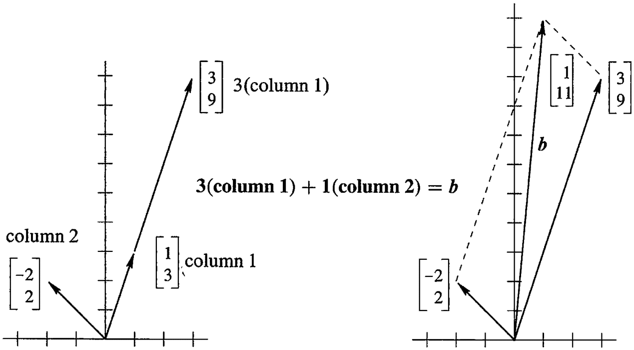
$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} x + \begin{pmatrix} -2 \\ 2 \end{pmatrix} y = \begin{pmatrix} 1 \\ 11 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} 3 + \begin{pmatrix} -2 \\ 2 \end{pmatrix} 1 = \begin{pmatrix} 1 \\ 11 \end{pmatrix}$$

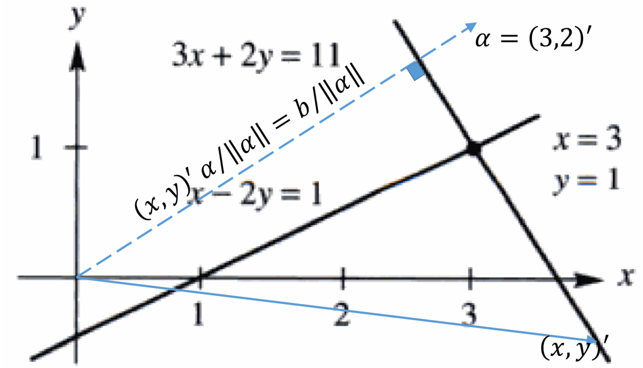
Row picture (vector equation of line intersection)

$$\begin{pmatrix} 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1$$

$$\begin{pmatrix} 3 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 11$$



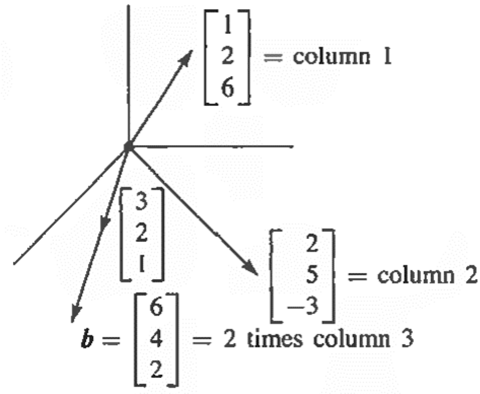
Projection $= (x, y)(\alpha/\|\alpha\|) = b/\|\alpha\| = 11/\sqrt{13}$.

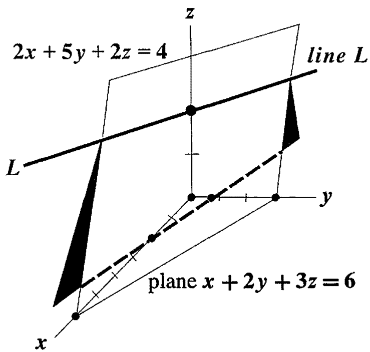


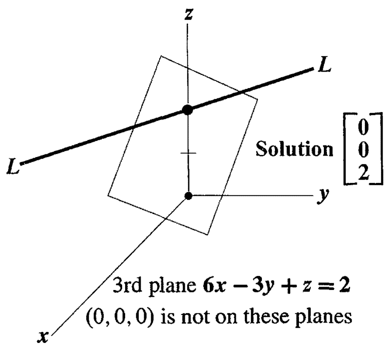
2.1.1 Three Equations in Three Unknowns

$$\begin{matrix} x+2y+3z=6 \\ 2x+5y+2z=4 \\ 6x-3y+z=2 \end{matrix} \quad\equiv\quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 6 \\ 4 \\ 2 \end{pmatrix} \quad\equiv\quad A\mathbf{x}=b \quad\equiv\quad C_1x+C_2y+C_3z=\begin{pmatrix} R_1\mathbf{x} \\ R_2\mathbf{x} \\ R_3\mathbf{x} \end{pmatrix}=b,$$

where $\mathbf{x}=(x,y,z)'$.







2.2 The Idea of Elimination

Systematic way to solve linear equations

- Find the **pivot** (1 in this example)
- Form the uper triangle system of equations.
- **Backsubstitution**.

Example 12 (2 equations)

$$\begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \qquad \text{(Before)}$$

$$\begin{aligned} x - 2y &= 1 \\ 8y &= 8 \end{aligned} \qquad \text{(After)}$$

Example 13 (3 equations)

$$\underset{\textit{Step 0}}{\begin{aligned} 2x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned}} \qquad \underset{\textit{Step 1}}{\begin{aligned} 2x + 4y - 2z &= 2 \\ 1y + 1z &= 4 \\ 1y + 5z &= 12 \end{aligned}} \qquad \underset{\textit{Step 2}}{\begin{aligned} \mathbf{2}x + 4y - 2z &= 2 \\ \mathbf{1}y + 1z &= 4 \\ \mathbf{4}z &= 8 \end{aligned}}$$

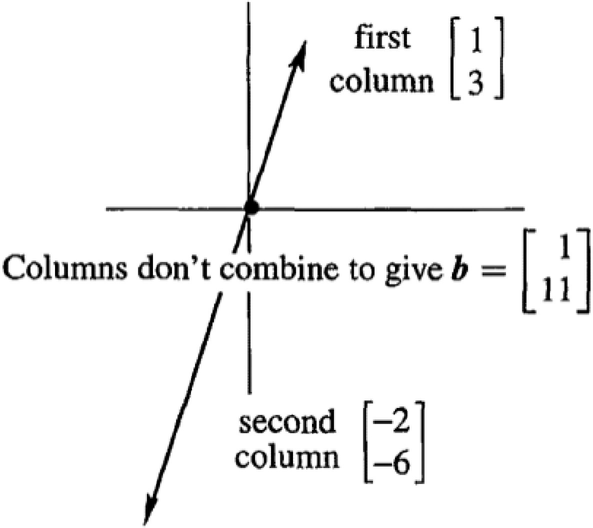
The solution is: $z = 2,\ y = 2,\ x = -1;$ *i.e.,* $(-1, 2, 4)$

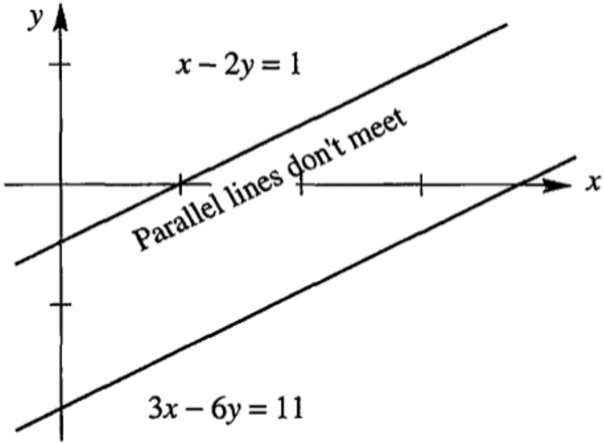
Failure 1: no solution

$$\begin{aligned} x-2y &= 1 \\ 3x-6y &= 11 \end{aligned} \qquad \text{(Before)}$$

$$\begin{aligned} x-2y &= 1 \\ 0y &= 8 \end{aligned} \qquad \text{(After)}$$

Interpretation from two perspectives



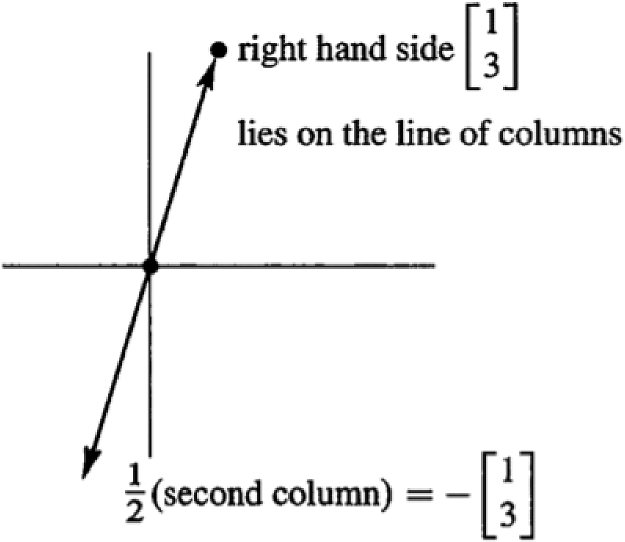


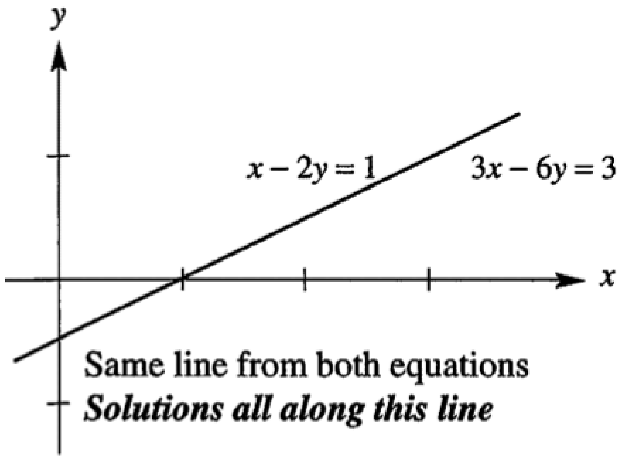
Failure 2: infinite solutions

$$\begin{aligned} x-2y &= 1 \\ 3x-6y &= 3 \end{aligned} \qquad \text{(Before)}$$

$$\begin{aligned} x-2y &= 1 \\ 0y &= 0 \end{aligned} \qquad \text{(After)}$$

Interpretation from two perspectives





2.3 Rules for Matrix Operations

Definition 14 (Matrix) *: A matrix $A_{m\times n}$ is a square array (of size $m \times n$) of "objects" (could be numbers could be other blocks of matrices). The element a_{ij} is located in row i and column j respectively. We say $A = (a_{ij})$ or in some books $A = ((a_{ij}))$ to denote:*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

- Languages handle matrices differently; e.g., Matlab images, C (row-wise), Fortran (column-wise), etc.
- Traversing matrices is $\Theta(m \times n)$.

2.3.1 Matrix Transpose

Definition 15 *The transpose of the matrix $A_{m\times n}$ is $\left(A'\right)_{n\times m}$, where $A_{ij} = \left(A'\right)_{ji}$*

Example 16

$$A = \begin{pmatrix} 18 & 17 & 11 \\ 19 & -4 & 0 \end{pmatrix},\ A' = \begin{pmatrix} 18 & 19 \\ 17 & -4 \\ 11 & 0 \end{pmatrix}$$

Notice:

- $\left(A'\right)' = A$.
- For vectors:
$$x = \begin{pmatrix} 19 \\ -4 \\ 0 \end{pmatrix},\ x' = \begin{pmatrix} 19 & -4 & 0 \end{pmatrix}.$$
we usually write $x = \begin{pmatrix} 19 & -4 & 0 \end{pmatrix}'$, or $x' = \begin{pmatrix} 19 & -4 & 0 \end{pmatrix}$ to save vertical space.

Definition 17 (Symmetric Matrices (around diagonal)) *A square matrix $A_{m\times m}$ is called symmetric if $A_{ij} = A_{ji}$; i.e., $A = A'$.*

Example 18 (write a SW to check the symmetry of) : $A = \begin{pmatrix} 18 & 17 & 11 \\ 17 & -4 & 0 \\ 11 & 0 & 2 \end{pmatrix}$

2.3.2 Matrix Partitioning

Definition 19 *A matrix $A_{p\times q}$ is said to be partitioned into $r \le p$ rows and $c \le q$ columns if it is written in the form*

$$
A = \begin{array}{c}
\begin{pmatrix}
A_{11} & A_{12} & \cdots & A_{1c} \\
A_{21} & A_{22} & \cdots & A_{2c} \\
\vdots & \vdots & & \vdots \\
A_{r1} & A_{r2} & \cdots & A_{rc}
\end{pmatrix}
\begin{array}{l}
\updownarrow p_1 \\
\updownarrow p_2 \\
\vdots \\
\updownarrow p_r
\end{array} \\
\begin{array}{cccc}
\overset{\longleftrightarrow}{q_1} & \overset{\longleftrightarrow}{q_2} & \cdots & \overset{\longleftrightarrow}{q_c}
\end{array}
\end{array}
\tag{2.1}
$$

where the block (submatrix) A_{ij} is a matrix $\left(A_{ij}\right)_{p_i q_j}$, and of course

$$
\begin{aligned}
A_{ij} &= (A_{ij})_{p_i \times q_j} \\
\sum_{i=1}^{r} p_i &= p \\
\sum_{j=1}^{c} q_j &= q.
\end{aligned}
$$

Example 20

$$
A = \begin{pmatrix}
1 & 6 & 8 & 9 & 3 & 8 \\
2 & 4 & 1 & 6 & 1 & 1 \\
3 & 3 & 6 & 1 & 2 & 1 \\
9 & 1 & 4 & 6 & 8 & 7 \\
6 & 8 & 1 & 4 & 3 & 2
\end{pmatrix}
= \left(\begin{array}{cccc|cc}
1 & 6 & 8 & 9 & 3 & 8 \\
2 & 4 & 1 & 6 & 1 & 1 \\
3 & 3 & 6 & 1 & 2 & 1 \\
\hline
9 & 1 & 4 & 6 & 8 & 7 \\
6 & 8 & 1 & 4 & 3 & 2
\end{array}\right)
$$

$$
= \begin{pmatrix}
(A_{11})_{3\times 4} & (A_{12})_{3\times 2} \\
(A_{21})_{2\times 4} & (A_{22})_{2\times 2}
\end{pmatrix}
$$

From the definition, this partitioning is not allowed:

$$
\left(\begin{array}{cc|cc|cc}
1 & 6 & 8 & \multicolumn{1}{c}{9} & 3 & 8 \\
\cline{3-6}
2 & 4 & 1 & 6 & 1 & 1 \\
3 & 3 & 6 & 1 & 2 & 1 \\
\hline
9 & \multicolumn{1}{c}{1} & 4 & 6 & 8 & 7 \\
6 & \multicolumn{1}{c}{8} & 1 & 4 & 3 & 2
\end{array}\right)
$$

Application: dividing vectors in regression.

Transposing Partitioned Matrix

It is quite easy to show that, the transpose of a partitioned matrix (2.1) is given by

$$A' = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1c} \\ A_{21} & A_{22} & \cdots & A_{2c} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rc} \end{pmatrix}' = \begin{pmatrix} A'_{11} & A'_{21} & \cdots & A'_{r1} \\ A'_{12} & A'_{22} & \cdots & A'_{r2} \\ \vdots & \vdots & & \vdots \\ A'_{1c} & A'_{2c} & \cdots & A'_{rc} \end{pmatrix}$$

Example 21

$$A = \left(\begin{array}{c|cc} 2 & 8 & 9 \\ 3 & 7 & 4 \end{array}\right) = \begin{pmatrix} A_{11} & A_{12} \end{pmatrix}$$

$$A' = \begin{pmatrix} A'_{11} \\ A'_{12} \end{pmatrix} = \left(\begin{array}{cc} 2 & 3 \\ \hline 8 & 7 \\ 9 & 4 \end{array}\right)$$

Paritioning into Vectors

Suppose that a_j is the jth column of $A_{r\times c}$. Then

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_c \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1c} \end{pmatrix},$$

where each submatrix is just a $r \times 1$ vector.

Similarly, it can be partitioned into r rows where α_i' is the ith row:

$$A = \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{r1} \end{pmatrix} = \begin{pmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_r' \end{pmatrix}$$

2.3.3 Matrix Trace

Definition 22 *For a square matrix $A_{m\times m}$, the trace,* $\mathrm{trace}\,(A)$*, (for short* $\mathrm{tr}\,(A)$*), is defined as the sum of diagonal elements; i.e.,*

$$\mathrm{tr}\,(A) = \sum_{i=1}^{m} A_{ii}.$$

HW: *write a C function to calculate the trace. (of course $\Theta(m)$)*

Corollary 23

$$\mathrm{tr}\,(A) = \mathrm{tr}\left(A'\right).$$

$$\mathrm{tr}\,(x) = x\ \forall x \in \mathbb{R}.$$

Proof.

$$\mathrm{tr}(A) = \sum_i A_{ii} = \sum_i (A')_{ii} = \mathrm{tr(A')}.$$

■

Example 24

$$A = \begin{pmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{pmatrix} \Longrightarrow \mathrm{tr}\,(A) = -4.$$

2.3.4 Addition, Subtraction, and Scaling

Definition 25 *For equal size matrices $A_{m\times n}$ and $B_{m\times n}$, and for a scalar λ:*

- *the matrix $C = A \pm B$ is defined as*
$$C_{ij} = A_{ij} \pm B_{ij},$$
- *the matrix $D = \lambda A$ is defined as*
$$D_{ij} = \lambda A_{ij},$$
- *we say that $A = B$ if $A_{ij} = B_{ij}\ \forall i, j$.*
- *and a matrix, all of whose components are zeros, is written as $\mathbf{0}_{m\times n}$.*
- *Of course, $A + \mathbf{0} = A$*

Corollary 26 *It is quite easy to show that*

$$(A+B)' = A' + B'$$
$$\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$$

Proof. Show that the general element ij of LHS equal to that of RHS.

$$\left((A+B)'\right)_{ij} = (A+B)_{ji} = A_{ji} + B_{ji} = (A')_{ij} + (B')_{ij} = (A'+B')_{ij}.$$
$$\operatorname{tr}(A+B) = \sum_i (A+B)_{ii} = \sum_i (A_{ii} + B_{ii}) = \sum_i A_{ii} + \sum_i B_{ii} = \operatorname{tr}(A) + \operatorname{tr}(B).$$

■

2.3.5 Matrix Multiplication

$$C = A_{m\times n}B_{n\times p} = \begin{pmatrix} a_{11} & & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & & b_{np} \end{pmatrix} = C_{m\times p}$$

The general element C_{ij} is the dot product of Row_i and Col_j:

$$C_{ik} = a_i' b_k = \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}.$$

However, we can partition either (or both) $A_{m\times n}$ and $B_{n\times p}$ as rows and/or columns to see the multiplication differently. This has a great value in mathematical treatments and semantics. We have only 4 ways to do that:

1. $A_{m\times 1}$, $B_{1\times p}$.
2. $A_{1\times n}$, $B_{n\times p}$.
3. $A_{1\times n}$, $B_{n\times 1}$.
4. $A_{m\times n}$, $B_{n\times 1}$.

Now, we will treat each case in detail.

1- As dot products

$$C = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} \begin{pmatrix} b_1 & \cdots & b_p \end{pmatrix} \qquad (A_{m\times 1} B_{1\times p} \text{ partitioning})$$

$$= \begin{pmatrix} a_1' b_1 & & a_1' b_p \\ \vdots & \ddots & \vdots \\ a_m' b_1 & & a_m' b_p \end{pmatrix} \begin{pmatrix} \sum_{j=1}^n a_{1j} b_{j1} & & \sum_{j=1}^n a_{1j} b_{jp} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^n a_{mj} b_{j1} & & \sum_{j=1}^n a_{mj} b_{jp} \end{pmatrix}$$

$$C_{ik} = a_i' b_k = \sum_{j=1}^n a_{ij} b_{jk} = a_{i1} b_{1k} + \cdots + a_{in} b_{nk}$$

Example 27

$$= \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \\ 1 & 4 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 3+4 & 2+16 & 0-4 \\ 3+5 & 2+20 & 0-5 \end{pmatrix}$$

$$= \begin{pmatrix} 7 & 18 & -4 \\ 8 & 22 & -5 \end{pmatrix}$$

2- As linear combinations of columns of A

$$
\begin{aligned}
C &= \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_{11} & & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & & b_{np} \end{pmatrix} \qquad (A_{1\times n} B_{n\times p} \text{ partitioning}) \\
&= \begin{pmatrix} b_{11}a_1 + \cdots + b_{n1}a_n & \cdots & b_{1p}a_1 + \cdots + b_{np}a_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_j b_{j1}a_j & \cdots & \sum_j b_{jp}a_j \end{pmatrix} \\
&= \begin{pmatrix} c_1 & \cdots & c_p \end{pmatrix} \\
C_{ik} &= (c_k)_i = \Big(\sum_j b_{jk}a_j\Big)_i = \sum_j (b_{jk}a_j)_i = \sum_j b_{jk}a_{ij}.
\end{aligned}
$$

Example 28

$$
\begin{aligned}
C &= \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \\ 1 & 4 & -1 \end{pmatrix} \\
&= \left(3\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1\begin{pmatrix} 4 \\ 5 \end{pmatrix} \quad 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 4\begin{pmatrix} 4 \\ 5 \end{pmatrix} \quad 0\begin{pmatrix} 1 \\ 1 \end{pmatrix} + -1\begin{pmatrix} 4 \\ 5 \end{pmatrix} \right) \\
&= \left(\begin{pmatrix} 3+4 \\ 3+5 \end{pmatrix} \quad \begin{pmatrix} 2+16 \\ 2+20 \end{pmatrix} \quad \begin{pmatrix} 0-4 \\ 0-5 \end{pmatrix} \right) = \begin{pmatrix} 7 & 18 & -4 \\ 8 & 22 & -5 \end{pmatrix}
\end{aligned}
$$

3- As linear combinations of rows of B

$$C = \begin{pmatrix} a_{11} & & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & & a_{mn} \end{pmatrix} \begin{pmatrix} b'_1 \\ \vdots \\ b'_n \end{pmatrix} \quad (A_{m\times n} B_{n\times 1} \text{ partitioning})$$

$$= \begin{pmatrix} a_{11} b'_1 + \cdots + a_{1n} b'_n \\ \vdots \\ a_{m1} b'_1 + \cdots + a_{nm} b'_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^n a_{1j} b'_j \\ \vdots \\ \sum_{j=1}^n a_{mj} b'_j \end{pmatrix}$$

$$= \begin{pmatrix} c'_1 \\ \vdots \\ c'_m \end{pmatrix}$$

$$C_{ik} = (c'_i)_k = \big(\sum_j a_{ij} b'_j\big)_k = \sum_j \big(a_{ij} b'_j\big)_k = \sum_j a_{ij} b_{jk}.$$

Example 29

$$= \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \\ 1 & 4 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 1\begin{pmatrix} 3 & 2 & 0 \end{pmatrix} + 4\begin{pmatrix} 1 & 4 & -1 \end{pmatrix} \\ 1\begin{pmatrix} 3 & 2 & 0 \end{pmatrix} + 5\begin{pmatrix} 1 & 4 & -1 \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{pmatrix} 3+4 & 2+16 & 0-4 \end{pmatrix} \\ \begin{pmatrix} 3+5 & 2+20 & 0-5 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 7 & 18 & -4 \\ 8 & 22 & -5 \end{pmatrix}$$

4- As summation of outer products, each is a matrix

$$C = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1' \\ \vdots \\ b_n' \end{pmatrix} \quad (A_{1\times n} B_{n\times 1} \text{ partitioning})$$

$$= a_1 b_1' + \cdots + a_n b_n'$$

$$= \sum_{j=1}^{n} a_j b_j',$$

$$a_j b_j' = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} b_j' = \begin{pmatrix} a_{1j} b_j' \\ \vdots \\ a_{nj} b_j' \end{pmatrix},$$

$$C_{ik} = \Big(\sum_j a_j b_j'\Big)_{ik} = \sum_j \big(a_j b_j'\big)_{ik} = \sum_j a_{ij} b_{jk}.$$

Example 30

$$= \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \\ 1 & 4 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \end{pmatrix} \begin{pmatrix} 1 & 4 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 2 & 0 \\ 3 & 2 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 16 & -4 \\ 5 & 20 & -5 \end{pmatrix} = \begin{pmatrix} 7 & 18 & -4 \\ 8 & 22 & -5 \end{pmatrix}$$

Partitioned Matrices and Multiplication (general case)

Subdevide each matrix to conforming number of blocks, e.g.

$$
\begin{aligned}
AB &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix} \\
&= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{pmatrix} && \text{(Must Conform)}
\end{aligned}
$$

In general: $A_{m\times n} B_{n\times p}$

$$
= \begin{array}{cccc} \overset{\longleftrightarrow}{n_1} & \overset{\longleftrightarrow}{n_2} & \cdots & \overset{\longleftrightarrow}{n_c} \\ \end{array}
$$

$$
= \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1c} \\ A_{21} & A_{22} & \cdots & A_{2c} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rc} \end{pmatrix}_{r\times c}
\qquad
\begin{array}{c} n_1 \updownarrow \\ n_2 \updownarrow \\ \vdots \\ n_c \updownarrow \end{array}
\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ B_{21} & B_{22} & \cdots & B_{2k} \\ \vdots & \vdots & & \vdots \\ B_{c1} & B_{c2} & \cdots & B_{ck} \end{pmatrix}_{c\times k}
$$

$n_1 + \cdots + n_c = n.$

Product with Diagonal Matrix

Definition 31 *A matrix D is diagonal if $D_{ij} = 0\ \forall i \neq j$; i.e.,*

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_m \end{pmatrix}.$$

Since there is no confusion, we subscript d_i instead of d_{ii}. We also, for short, write $D = \operatorname{diag}(d_1, \ldots, d_n)$

Row scaling:

$$D_{m\times m} A_{m\times n} = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_m \end{pmatrix} \begin{pmatrix} a'_1 \\ \vdots \\ a'_m \end{pmatrix} = \begin{pmatrix} d_1 a'_1 \\ \vdots \\ d_m a'_m \end{pmatrix}$$

Column scaling:

$$A_{m\times n} D_{n\times n} = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_n \end{pmatrix} = \begin{pmatrix} a_1 d_1 & \cdots & a_n d_n \end{pmatrix}$$

Definition 32 *The identity matrix I is a special case diagonal matrix and defined as*

$$I_{m\times m} = \operatorname{diag}(1, \ldots, 1)$$

It is obvious that $IA = AI = A$.

Transpose of a Product

Lemma 33 *For conforming matrices* $A_{m\times n}$ *and* $B_{n\times p}$*,*

$$(AB)' = B'A',$$

and more general

$$(A_1 \cdots A_n)' = A_n' \cdots A_1'.$$

Proof. The general element AB_{ik} is given by

$$(AB)_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk} = \sum_{j=1}^{n} \left(A'\right)_{ji} \left(B'\right)_{kj} = \sum_{j=1}^{n} \left(B'\right)_{kj} \left(A'\right)_{ji} = \left(B'A'\right)_{ki}.$$

Proving the second part is immediate by induction. ∎

Trace of a Product

The trace is defined only for a square matrix; hence, for a product to have a trace it must be $A_{m\times n}B_{n\times m}$.

Lemma 34 *For two-side conforming matrices $A_{m\times n}$ and $B_{n\times m}$,*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA),$$

and more general

$$\operatorname{tr}(A_1 \cdots A_n) = \operatorname{tr}(A_n \cdots A_1).$$

Proof. $A_{m\times n}B_{n\times m} = C_{m\times m}$, $B_{n\times m}A_{m\times n} = D_{n\times n}$:

$$\operatorname{tr}(AB) = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \operatorname{tr}(BA).$$

∎

Remark 1 *From the proof above, we see that*

$$\operatorname{tr}(AB) = \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m}\left(B'\right)_{ij} A_{ij},$$

i.e., it is the sum of products of each element of A multiplied by the corresponding element of B'. And if $B = A'$

$$\operatorname{tr}\left(AA'\right) = \operatorname{tr}\left(A'A\right) = \sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}A_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}^2,$$

i.e., it is the sum squares of all elements.

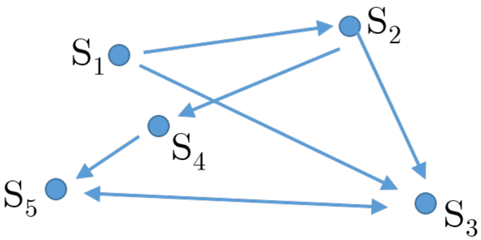
Example 35

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & 2 & 3 \\ -4 & 3 & 0 \end{pmatrix}, \\
\operatorname{tr}\left(AA'\right) &= 1^2 + 2^2 + 3^2 + (-4)^2 + 3^2 + 0^2 = 39
\end{aligned}
$$

Power of a Matrix

$$A^k = AA\cdots A,\ k \text{ times} \qquad (A \text{ must be square; why?})$$

Example 36 (Graph Theory) :



- *The traffic is represented as a matrix T, where a path from S_i to S_j exists if $T_{ij} = 1$.*

$$T = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

- *Number of ways of getting from S_i to S_k in exactly 2 steps is $\sum_j T_{ij} T_{jk} = \left(T^2\right)_{ik}$.*
- *Number of ways of getting from S_i to S_k in exactly 3 steps is $\sum_j T_{ij} \left(T^2\right)_{jk} = \left(T^3\right)_{ik}$.*

$$T^2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, T^3 = \begin{pmatrix} 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

- *Number of ways of getting from S_i to S_k in exactly r steps is $\sum_j T_{ij} \left(T^{r-1}\right)_{jk} = (T^r)_{ik}$.*
- *There is no path from S_i to S_k only if $\sum_{r=1}^{\infty} (T^r)_{ik} = 0$.*
- *What is $\sum_{r=1}^{\infty} T^r$?*

2.3.6 The Laws of Algebra

Theorem 37 *$\forall A_{m\times n}$, $B_{m\times n}$, $C_{m\times n}$, c scalar, we have*

$$A+B=B+A \qquad \text{(commulative)}$$
$$c\,(A+B)=cA+cB \qquad \text{(distributive)}$$
$$A+(B+C)=(A+B)+C, \qquad \text{(associative)}$$

and

$$C\,(A+B)=CA+CB, \qquad (\forall A_{m\times n}, B_{m\times n}, C_{k\times m})$$
$$(A+B)\,C=AC+BC, \qquad (\forall A_{m\times n}, B_{m\times n}, C_{n\times p})$$
$$A\,(BC)=(AB)\,C \qquad (\forall A_{m\times n}, B_{n\times p}, C_{p\times q})$$
$$A_{m\times n}B_{n\times m}\neq B_{n\times m}A_{m\times n}$$
$$A_{m\times m}B_{m\times m}=\neq B_{m\times m}A_{m\times m}$$

Example 38 (Counter Example for $AB\neq BA$)

$$\begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}0&2\\3&3\end{pmatrix}=\begin{pmatrix}6&8\\12&18\end{pmatrix}=\begin{pmatrix}0&2\\3&3\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$$

$$\underbrace{\begin{pmatrix}1&2\\3&4\end{pmatrix}}_{A}\underbrace{\begin{pmatrix}0&1\\3&5\end{pmatrix}}_{B}=\begin{pmatrix}6&11\\12&23\end{pmatrix}\neq\begin{pmatrix}3&4\\18&26\end{pmatrix}=\underbrace{\begin{pmatrix}0&1\\3&5\end{pmatrix}}_{B}\underbrace{\begin{pmatrix}1&2\\3&4\end{pmatrix}}_{A}$$

Proof of multiplication associative rule. $A_{m\times n}B_{n\times p}C_{p\times q}$

$$
\begin{aligned}
(AB)_{ij} &= \sum_{r=1}^{n} A_{ir}B_{rj}\\
((AB)\,C)_{ik} &= \sum_{j=1}^{p} (AB)_{ij}\,C_{jk}\\
&= \sum_{j=1}^{p}\sum_{r=1}^{n} A_{ir}B_{rj}C_{jk}\\
&= \sum_{r=1}^{n} A_{ir}\sum_{j=1}^{p} B_{rj}C_{jk}\\
&= \sum_{r=1}^{n} A_{ir}\,(BC)_{rk}\\
&= (A\,(BC))_{ik}
\end{aligned}
$$

■

Example 39 *Factor* $Y = XPX + QX^2 + X$ *and find the constraints on the order of matrices. It is clear that all matrices will be of order* $m \times m$.

$$\begin{aligned} Y &= XPX + QX^2 + X \\ &= (XP + QX + I)\, X \end{aligned}$$

$$XPX + QXX + X$$

Product with Scalar and Quadratic Forms

Back to Definition 2 it is very important, sometimes, to make sure of conforming even for scalars; i.e., we write

$$y_{m\times 1} a_{1\times 1} \textbf{ NOT } ay.$$

This is because, sometimes, $a_{1\times 1}$ itself is a matrix multiplication that if dissembled it should conform with the remaining of equation

$$\begin{aligned} a_{1\times 1} &= x'_{1\times m} A_{m\times m} x_{m\times 1} \\ y_{n\times 1} a_{1\times 1} &= \underline{y_{n\times 1} x'_{1\times m}} A_{m\times m} x_{m\times 1} \\ a_{1\times 1} y_{n\times 1} &= x'_{1\times m} A_{m\times m} \underline{x_{m\times 1} y_{n\times 1}} && \text{(WRONG!)} \end{aligned}$$

Example 40 (Quadratic Form) *For any square matrix A, the form $y_{1\times 1} = x'_{1\times n} A_{n\times n} x_{n\times 1}$ is called quadratic form; it contains all quadratic and bilinear terms. (For scalar case, simply it is $y = xax = ax^2$).*

$$
\begin{aligned}
y &= \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_i x_i a_{i1} & \sum_i x_i a_{i2} & \dots & \sum_i x_i a_{in} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
&= \sum_j \left(\sum_i x_i a_{ij} \right) x_j \\
&= \sum_j \sum_i a_{ij} x_j x_i \\
&= \sum_i a_{ii} x_i^2 + \sum\sum_{i\neq j} \underbrace{a_{ij}}_{\text{all off diagonal}} x_i x_j \\
&= \sum_i a_{ii} x_i^2 + \sum\sum_{i>j} \left(\underbrace{a_{ij}}_{LT} + \underbrace{a_{ji}}_{UT} \right) x_i x_j,
\end{aligned}
$$

this is because, e.g., $a_{13}x_1x_3 + a_{31}x_3x_1 = (a_{13} + a_{31})\, x_1 x_3$.

Complexity of Ver. 1: *we sum $\left(n^2 - n\right)$ off-diagonal terms, each term is 2 multiplications $\left(a_{ij}x_i x_j\right)$; therefore* *total steps is given by:*

$$\begin{aligned}\#\, of\ steps &= \left(n^2-n\right)(2M) + \left(n^2-n-1\right)(S)\\ &= 2\left(n^2-n\right)M + \left(n^2-n-1\right)S\end{aligned}$$

Complexity of Ver. 2:, *we sum* $\left(n^2-n\right)/2$ *lower triangular term, each term is one summation and 2 multiplications; therefore*

$$\begin{aligned}\#\, of\ steps &= \frac{\left(n^2-n\right)}{2}(2M+1S) + \left(\frac{\left(n^2-n\right)}{2}-1\right)S\\ &= \left(n^2-n\right)M + \left(n^2-n-1\right)S.\end{aligned}$$

Ver. 2 is half the number of multiplications of Ver. 1; it is almost double speed gained by a simple trick.

For the following quadratic form y, expand column wise $\left(\sum_j \sum_i\right)$:

$$\begin{aligned}y &= \begin{pmatrix}x_1 & x_2 & x_3\end{pmatrix}\begin{pmatrix}1 & 2 & 3\\ 4 & 7 & 6\\ 2 & -2 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}\\ &= x_1^2 + 4x_2x_1 + 2x_3x_1 + 2x_1x_2 + 7x_2^2 - 2x_3x_2 + 3x_1x_3 + 6x_2x_3\\ &= x_1^2 + (2+4)\,x_1x_2 + (3+2)\,x_1x_3 + 7x_2^2 + (6-2)\,x_2x_3\\ &= x_1^2 + 6x_1x_2 + 5x_1x_3 + 7x_2^2 + 4x_2x_3.\end{aligned}$$

Without expansion, it is obvious that, e.g.,

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 5 & 7 & 3 \\ 4 & 1 & 0 \end{pmatrix}}_{B} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

because $\left(A_{ij} + A_{ji}\right) = \left(B_{ij} + B_{ji}\right) \ \forall i, j.$

Hence, we can replace the matrix A *in any quadratic form* $y = x'Ax$ *by a symmetric matrix* $\Sigma = \left(A + A'\right)/2$ *whose diagonals and off-diagonals have:*

$$\begin{aligned} \sigma_{ii} &= (a_{ii} + a_{ii})/2 = a_{ii} \\ \left(\sigma_{ij} + \sigma_{ji}\right) &= \left(a_{ij} + a_{ji}\right)/2 + \left(a_{ji} + a_{ij}\right)/2 \\ &= a_{ij} + a_{ji} \\ x'Ax &= x'\Sigma x \end{aligned}$$

Example 41 *Expand and simplify $y = (x-\mu)' \Sigma (x-\mu)$, where x and μ are vectors and Σ is a symmetric matrix.*

$$
\begin{aligned}
y &= (x-\mu)' \Sigma (x-\mu) \\
&= (x' - \mu') \Sigma (x-\mu) \\
&= x'\Sigma x - x'\Sigma\mu - \mu'\Sigma x + \mu'\Sigma\mu \\
&= x'\Sigma x - x'\Sigma\mu - \left(\mu'_{1\times p}\Sigma_{p\times p} x_{p\times 1}\right)' + \mu'\Sigma\mu && (scalar' = scalar) \\
&= x'\Sigma x - x'\Sigma\mu - x'\Sigma\mu + \mu'\Sigma\mu \\
&= x'\Sigma x - 2x'\Sigma\mu + \mu'\Sigma\mu
\end{aligned}
$$

2.4 Elimination Using Matrices

Back to the linear system of equations (Ex., 13) (and using **pivots** for elemination):

$$\begin{aligned} 2x+4y-2z&=2\\ 4x+9y-3z&=8\\ -2x-3y+7z&=10 \end{aligned} \implies \begin{pmatrix} 2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7 \end{pmatrix}\begin{pmatrix} x\\ y\\ z \end{pmatrix}=\begin{pmatrix} 2\\ 8\\ 10 \end{pmatrix}$$

$$(\text{col1})\,x+(\text{col2})\,y+(\text{col3})\,z=\begin{pmatrix} (\text{row1})\,.x\\ (\text{row2})\,.y\\ (\text{row3})\,.z \end{pmatrix}=b$$

To eliminate: $R_2^{\text{new}}=R_2+(-2)\times R_1$, which can be accomplished by the matrix multiplication:

$$\begin{pmatrix} 1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7 \end{pmatrix}\begin{pmatrix} x\\ y\\ z \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2\\ 8\\ 10 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 4 & -2\\ 0 & 1 & 1\\ -2 & -3 & 7 \end{pmatrix}\begin{pmatrix} x\\ y\\ z \end{pmatrix}=\begin{pmatrix} 2\\ 4\\ 10 \end{pmatrix}.$$

We denote the elimination matrix $\begin{pmatrix} 1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}$ by $E_{21}(-2)$.

Definition 42 *The elimination matrix $E_{ij}(l)$ is an identity matrix except the element $e_{ij}=l$ to perform: $R_i^{new}=R_i+l\times R_j$. If not ambiguous we write E_{ij}*

$$E_{31}(1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 10 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 12 \end{pmatrix}$$

Finally,

$$E_{32}(-1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix},$$

whose solution is $z = 2,\ y = 2,\ x = -1$.
The summary of that is

$$E_{32}(-1)E_{31}(1)E_{21}(-2)AX = E_{32}(-1)E_{31}(1)E_{21}(-2)b.$$

Just for simpler notation (with same everything), we could have made up the augmented matrix

$$(A|\ b) = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{array}\right)$$

,then

$$E_{32}E_{31}E_{21}\,(A|\ b)$$

Definition 43 *The permutation matrix P_{ij} is an identity matrix except that in Rows i and j (to be permuted) the ones are located in p_{ij}, p_{ji} respectively; e.g.,*

$$P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Of course, $P_{ij} = P_{ji}$.

This is needed to swap equations when the pivot is zero.

Example 44

$$\begin{aligned} x+2y+2z&=1\\ 4x+8y-9z&=3\\ 3y+2z&=1 \end{aligned} \implies (A|\,b)=\left(\begin{array}{ccc|c} 1&2&2&1\\ 4&8&9&3\\ 0&3&2&1 \end{array}\right)$$

$$\begin{pmatrix} 1&0&0\\ -4&1&0\\ 0&0&1 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|r} 1&2&2&1\\ 0&0&1&-1\\ 0&3&2&1 \end{array}\right) \qquad (E_{21}(-4))$$

$$\begin{pmatrix} 1&0&0\\ 0&0&1\\ 0&1&0 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|r} 1&2&2&1\\ 0&3&2&1\\ 0&0&1&-1 \end{array}\right). \qquad (P_{23})$$

This is called Gauss elimination, and by back-substitution, $z=-1,\ y=1,\ x=1.$ *Jordan would go further to get pivots on diagonal and zeros elsewhere.*

$$= \left(\begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \qquad (E_{23}(-2))$$

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \qquad (E_{13}(-2))$$

$$\begin{pmatrix} 1 & -\frac{2}{3} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \qquad (E_{12}(\tfrac{-2}{3}))$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \swarrow = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right) \qquad (D(1, \tfrac{1}{3}, 1))$$

$$= (I|\ solution),$$

where the solution is : $x = 1, y = 1, z = -1$.
The summary of that is:

$$DE_{12}E_{13}E_{23}P_{23}E_{21}\,(A|\ b)$$

Solution of system or linear equations is nothing but multiplication by E***s,*** P***s, and finally*** D

Example 45 (Elimination by blocks) *Using the first pivot, we can eliminate all elements underneath using a single matrix. Write the matrix A as*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

In general we eliminate by:

$$(E_{n\ n-1})\dots(E_{n2}\dots E_{42}E_{32})\,(E_{n1}\dots E_{31}E_{21})$$

$$\begin{aligned}
E &= E_{31}E_{21} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -a_{31}/a_{11} & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ -a_{21}/a_{11} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ -a_{21}/a_{11} & 1 & 0 \\ -a_{31}/a_{11} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ -A_{21}/a_{11} & I \end{pmatrix}
\end{aligned}$$

The power of block treatment, allow us to write

$$\begin{aligned}
EA &= \begin{pmatrix} 1 & \mathbf{0} \\ -A_{21}/a_{11} & I \end{pmatrix}\begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & A_{12} \\ \mathbf{0} & A_{22} - A_{21}A_{12}/a_{11} \end{pmatrix}.
\end{aligned}$$

Of course A can be replaced by $(A|\ b)$. *For this example*

$$
\begin{aligned}
(A|\ b) &= \left(\begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 4 & 8 & 9 & 3 \\ 0 & 3 & 2 & 1 \end{array}\right) \\
A_{22} - A_{21}A_{12}/a_{11} &= \begin{pmatrix} 8 & 9 & 3 \\ 3 & 2 & 1 \end{pmatrix} - \begin{pmatrix} 4 \\ 0 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 \end{pmatrix} /1 \\
&= \begin{pmatrix} 8 & 9 & 3 \\ 3 & 2 & 1 \end{pmatrix} - \begin{pmatrix} 8 & 8 & 4 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 3 & 2 & 1 \end{pmatrix}
\end{aligned}
$$

This gives

$$
\left(\begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 3 & 2 & 1 \end{array}\right),
$$

which would be obtained of course having multiplied by $E_{21}(-4)$

2.5 Inverse Matrices

Definition 46 *The* ***square*** *matrix* $A_{p\times p}$ *is invertible if there exists a matrix* A^{-1} *such that*

$$A_l^{-1}A = AA_r^{-1} = I_{p\times p}$$

Hint: *we will show soon that* $A_l^{-1} = A_r^{-1} = A^{-1}$*. But we have to be cautious and* ***rigorous****, since* $AB \neq BA$ *in general.* ■

Motivation: for scalar a

$$\begin{aligned} aX &= b \\ a^{-1}aX &= a^{-1}b \\ 1\,X &= a^{-1}b \\ X &= a^{-1}b \end{aligned}$$

Analogously, what is A^{-1} such that

$$\begin{aligned} AX &= b \\ A^{-1}AX &= A^{-1}b \\ IX &= A^{-1}b \\ X &= A^{-1}b, \end{aligned}$$

although finding A^{-1} is more computational expensive than solving by elimination as we will see.

Lemma 47 *If both left and right inverses exist they are equal*

Proof. Suppose the left and right inverses of A are A_l^{-1} and A_r^{-1} (so that $A_l^{-1}A = AA_r^{-1} = I$); then consider $A_l^{-1}AA_r^{-1}$

$$A_r^{-1} = \left(A_l^{-1}A\right)A_r^{-1} = \underleftrightarrow{A_l^{-1}AA_r^{-1}} = A_l^{-1}\left(AA_r^{-1}\right) = A_l^{-1}$$

■

This Lemma is different from the last two statements in Lemma 51 (will be proven shortly), from which we can say:

1. If left (or right) inverse exists the right (or left) exists and equals it. Stated differently, if $AB = I$ then $BA = I$.

2. If the inverse exists it is unique. So, we cannot find $B_1A = I$ and $B_2A = I$ with $B_1 \neq B_2$

Therefore,
Either: the square matrix A has no inverse
Or: the left and right inverses are identical and unique.

Lemma 48 (inverse of special matrices) :

1. *Any* 2×2 *matrix:*
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

2. *Any* $n \times n$ *diagonal matrix:*
$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}^{-1} = \begin{pmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{pmatrix}$$

3. *Any pivot cancellation matrix:*
$$E_{ij}^{-1}(l) = E_{ij}(-l)$$

4. *Any permutation matrix:*
$$\left(P_{ij}\right)^{-1} = P_{ij}$$

Proof. The proof is by direct multiplication from both sides; it is obvious for 1 and 2. For 3,

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & -l & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & l & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & -l \times 1 + l & \ddots & \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Proving 4 follows exactly the same line. In few words, since P_{ij} is I with rows i and j swapped then $P_{ij}P_{ij}$ swaps again the same rows to bring it back to I. ■

Example 49 *Consider $E_{21}(-5)$, then*

$$E_{21}(-5)\,A = \begin{pmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 - 5R_1 \\ R_3 \end{pmatrix},$$

$$E_{21}(5)\,(E_{21}(-5)\,A) = \begin{pmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 - 5R_1 \\ R_3 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix},$$

i.e., it subtracts what E added. Of course, $EE^{-1} = E^{-1}E = I$.

Calculating A^{-1} by Gauss-Jordan Elimination

Consider $A_{n\times n}$ and its right inverse exists: $A_r^{-1} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$. Then,

$$A\begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} = I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix}$$

$$\begin{pmatrix} Ax_1 & \cdots & Ax_n \end{pmatrix} = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix}$$

$$Ax_1 = e_1$$

$$\vdots$$

$$Ax_n = e_n$$

Then, finding A^{-1} is nothing but solving by elimination n systems of equations, each is $n \times n$:

$$Ax_i = e_i, i = 1, \ldots, n.$$

Example 50 .

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Find A^{-1} *using the augmented matrix* $(A|\ I) =$

$$\begin{aligned}
&= \left(\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right) \\
&\to \left(\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right) && (E_{21}(\tfrac{1}{2})) \\
&\to \left(\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right) && (E_{32}(\tfrac{2}{3});\ \text{Gauss stops here}) \\
&\to \left(\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right) && (E_{23}(\tfrac{3}{4})) \\
&\to \left(\begin{array}{rrr|rrr} 2 & 0 & 0 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right) && (E_{12}(\tfrac{2}{3})) \\
&\to \left(\begin{array}{rrr|rrr} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right) && (D(\tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4})) \\
&= (I|\ A_r^{-1}) && \text{(reduced echelon form)}
\end{aligned}$$

In summary:
$DE_{12}E_{23}E_{32}E_{21}\,(A|\ I) = (I|\ A_r^{-1})$.

Lemma 51 (Connection between A^{-1} and pivots) *:*

1. *If A has n pivots A^{-1} exists and*
$$M = A_l^{-1} = A^{-1} = A_r^{-1} = X,$$
where M is the multiplication of the elemination matrices and X is the solution of $AX = I$.

2. *If either inverse exists A has n pivots and hence A^{-1} exists. (This means if $AB = I$ then $BA = I$)*

3. *If the inverse exists then it is unique along with pivots and the solution to $Ax = b$.*

Proof. 1: If the pivots exist then this has been produced to initially solve the problem $AX = I$, and the solution X went to the right side; therefore the solution X is A_r^{-1}. In parallel, the solution is nothing but a series of matrix multiplications:

$$\overleftarrow{D(E_{12})\dots(E_{1\ n-1}\dots E_{n-3\ n-1}E_{n-2\ n-1})(E_{1n}\dots E_{n-2\ n}E_{n-1\ n})}\cdot$$
$$\overleftarrow{(E_{n\ n-1})\dots(E_{n2}\dots E_{42}E_{32})(E_{n1}\dots E_{31}E_{21})}A = I,$$

in the form $MA = I$; hence, M is A_l^{-1}. Since both inverses exist, they are equal (Lemma 47).

2: If A_r^{-1} exists ($AX = I$**)** we will prove A has n pivots by contradiction. Assume that A does not have n pivots (the elimination matrices MA produces a matrix with zero row):

$$\text{zero row mat.} = \underbrace{(MA)}_{\text{zero row mat.}} X = \overleftrightarrow{MAX} = M(AX) = MI = M.$$

However, M cannot have a zero row; otherwise it would produce a zero row matrix; while **by construction**, it should produce n pivots not a zero row; a contradiction. Hence, A has n pivots and from 1 above $M = A_l^{-1} = A^{-1} = A_r^{-1} = X$ (which means $XA = I$).

If A_l^{-1} exists ($XA = I$), then $AX = I$ with which we have just started lines above.

3: Assume that A has two inverses A_1 and A_2, so that $A_1A = AA_1 = I$ and $A_2A = AA_2 = I$.

$$\begin{aligned} A_1A &= I \\ A_1AA_2 &= A_2 \\ A_1 &= A_2 \end{aligned}$$

Since the inverse is unique, the elimination process cannot produce different pivots; hence they are unique too and the solution to $Ax = b \; \forall b$ will be unique as well and equals to $A^{-1}b$. ■

Lemma 52 *If A is symmetric, then its inverse is symmetric.*

Proof. Suppose that B is an inverse then

$$\begin{aligned} BA &= I \\ A'B' &= I \\ AB' &= I \\ BAB' &= B \\ B' &= B. \end{aligned}$$

■

Lemma 53

1. *Suppose that A, B, are invertible,* $(AB)^{-1} = B^{-1}A^{-1}$.

2. *And in general* $(A_1 \cdots A_n)^{-1} = A_n^{-1} \cdots A_1^{-1}$.

Proof. For the first part,

$$
\begin{aligned}
(AB)^{-1}(AB) &= I \\
(AB)^{-1}(AB)B^{-1}A^{-1} &= B^{-1}A^{-1} \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(AB)(AB)^{-1} &= I \\
B^{-1}A^{-1}(AB)(AB)^{-1} &= B^{-1}A^{-1} \\
(AB)^{-1} &= B^{-1}A^{-1}.
\end{aligned}
$$

The proof of part 2 is immediate by induction. ■

Lemma 54 *If* $AX = \mathbf{0}$ *and* $X \neq \mathbf{0}$ *then A is not invertible.*

Proof. Given $X \neq \mathbf{0}$, suppose that A^{-1} exists;

$$
\begin{aligned}
AX &= \mathbf{0} \\
A^{-1}AX &= A^{-1}\mathbf{0} \\
X &= \mathbf{0},
\end{aligned}
$$

a contradiction; hence A^{-1} does not exist. ■

2.6 Elimination Using Matrices is $A = LU$ Factorization

Back to Sec. 2.4

$$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7\end{pmatrix} = \begin{pmatrix}2 & 4 & -2\\ 0 & 1 & 1\\ 0 & 0 & 4\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & 0\\ -2 & 1 & 0\\ 3 & -1 & 1\end{pmatrix}\begin{pmatrix}2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7\end{pmatrix} = \begin{pmatrix}2 & 4 & -2\\ 0 & 1 & 1\\ 0 & 0 & 4\end{pmatrix}$$

$$E_{32}(-1)E_{31}(1)E_{21}(-2)\ A = M_L\ A = U$$

$$\begin{aligned} A &= M_L^{-1}\ U = \big(E_{32}(-1)E_{31}(1)E_{21}(-2)\big)^{-1}\ U\\ &= \big(E_{21}(2)E_{31}(-1)E_{32}(1)\big)U\\ A &= L\ U \end{aligned}$$

$$\begin{aligned}\begin{pmatrix}2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7\end{pmatrix} &= \begin{pmatrix}1 & 0 & 0\\ 2 & 1 & 0\\ -1 & 1 & 1\end{pmatrix}\begin{pmatrix}2 & 4 & -2\\ 0 & 1 & 1\\ 0 & 0 & 4\end{pmatrix}\\ &= \begin{pmatrix}1 & 0 & 0\\ 2 & 1 & 0\\ -1 & 1 & 1\end{pmatrix}\begin{pmatrix}2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 4\end{pmatrix}\begin{pmatrix}1 & 2 & -1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}\end{aligned}$$

$$A = L\ D\ U.$$

Intuitively:

$$U_3 = A_3 - l_{31}U_1 - l_{32}U_2$$
$$A_3 = U_3 + l_{31}U_1 + l_{32}U_2$$

Remark: L stores the **Gauss-elimination steps** on A, which ends up to U.

Example 55 (Using $L\,U$ in solving equations:)

$$AX = b \qquad \equiv \qquad L\,(U\,X) = b$$

Then, solve $L\,C = b$ to find C, then solve $U\,X = C$ to find X.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix} \qquad \text{(Gauss-elimination for } b\text{)}$$

$$c_1 = 2 \qquad 2c_1 + c_2 = 8 \longrightarrow c_2 = 4 \qquad -c_1 + c_2 + c_3 = 10 \longrightarrow c_3 = 8$$

$$\begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix} \qquad \text{(same obtained with augmenting)}$$

$$4z = 8 \longrightarrow z = 2 \qquad y + z = 4 \longrightarrow y = 2 \qquad 2x + 4y - 2z = 2 \longrightarrow x = -1$$

Lemma 56 (**$A = L\,U$ factorization**) *: for the case of no permutation, we get*

$$\begin{aligned}
\left(E_{n\ n-1}\right)\cdots\left(E_{n2}\cdots E_{42}E_{32}\right)\left(E_{n1}\cdots E_{31}E_{21}\right)A &= U\\
M_L\ A &= U\\
A &= M_L^{-1}\ U\\
&= \left(E_{21}^{-1}E_{31}^{-1}\cdots E_{n1}^{-1}\right)\left(E_{32}^{-1}E_{42}^{-1}\cdots E_{n2}^{-1}\right)\cdots\left(E_{n\ n-1}^{-1}\right)U\\
&= L\,U,
\end{aligned}$$

where: (1) both M_L, $M_L^{-1}(= L)$ are LTMs, and (2) L has L_{ij} equals directly to the element of E_{ij} as opposed to M_L. The proof is immediate from the following two more general lemmas. ***Hint:*** *to prove that the elements of M_L are not directly the elements of E_{ij} a single counter example is enough.*

Lemma 57 *Multiplication of two lower (or upper) triangular matrices is a lower (or upper) triangular matrix. The diagonal will be one if $A_{ii}B_{ii} = 1$ ($A_{ii} = B_{ii} = 1$ is a special case).*

Proof. Suppose A, B are LTMs; i.e., $A_{ij} = B_{ij} = 0\ \forall i < j$. Then, the element $C_{ij}, i \leq j$ will be

$$C_{ij} = \sum_k A_{ik}B_{kj} = \sum_{k<i} A_{ik}B_{kj} + A_{ii}B_{ij} + \sum_{k>i} A_{ik}B_{kj} = \sum_{k<i} A_{ik}\,0 + A_{ii}B_{ij} + \sum_{k>i} 0\ B_{kj} = A_{ii}B_{ij}$$

which is 0 for $i < j$ and $A_{ii}B_{ii}$ for $i = j$. Hence, it is obvious that M_L is LTM with ones on the diagonal ■

Lemma 58 *Consider any two LTMs $A,\ B$ with the following properties*

$$
\begin{aligned}
A_{ij} &= B_{ij} = 0 && \forall i < j\\
A_{ii} &= B_{ii} = 1 && \\
A_j &= e_j && \forall j > J\\
B_j &= e_j && \forall j < J\\
A_{iJ} &= 0 && \forall I < i\\
B_{iJ} &= 0 && \forall J < i \leq I.
\end{aligned}
$$

Since $C_j = \sum_i B_{ij} A_i$, we get:

$$
\begin{aligned}
C_j &= \sum_{i \neq j} B_{ij} A_i + B_{jj} A_j = 0 + A_j = A_j, && (\forall j < J)\\
C_j &= \sum_{i} B_{ij} A_i = \sum_{i<j} B_{ij} A_i + \sum_{j \leq i} B_{ij} A_i = 0 + \sum_{j \leq i} B_{ij} e_i = B_j, && (\forall j > J)\\
C_J &= \sum_{i<J} B_{iJ} A_i + B_{JJ} A_J + \sum_{J<i\leq I} B_{iJ} A_i + \sum_{I<i} B_{iJ} A_i = 0 + A_J + 0 + \sum_{I<i} B_{iJ} e_i. && (j = J)
\end{aligned}
$$

Hence, each element of A and B goes to C directly in the same position.

Example 59 (Common Mistake:) *Do the elements of Es go directly to M_L and hence:*

$$\begin{aligned}
L^{-1}_{i,j} &= -L_{i,j}, && i > j\\
L^{-1}_{i,j} &= L_{i,j} = 1, && i = j\\
L^{-1}_{i,j} &= L_{i,j} = 0, && i < j.
\end{aligned}$$

Lemma 60 *If A has a row starting with zero, so does the same row in L; and when a column in A starts with zero, so does the same column in U*

Proof. If $A_{i1} = 0$, then $L_{i1} = 0$ is immediate from

$$0 = A_{i1} = \sum_k L_{ik} U_{k1} = L_{i1} U_{11} + \sum_{k>1} L_{ik} 0;$$

Also, it could be immediate from the fact that if a row in A has zero, it does not need elimination and hence the element of its E matrix will be zero. This saves computer time.
On the other hand, if $A_{1j} = 0$, then $U_{1j} = 0$ is immediate from

$$0 = A_{1j} = \sum_k L_{1k} U_{kj} = 1 U_{1j} + \sum_{k>1} 0 U_{kj},$$

which completes the proof. ∎

2.7 Computational Issues: Scientific Computing Environments (SCEs), Examples, and Complexity

2.7.1 On Scientific Computing Environments and Libraries

EISPACK : early 1970s, for solving symmetric, unsymmetric, and generalized eigenproblems.

LINPACK : late 1970s for solving linear equations and least squares problems.

BLAS (**B**asic **L**inear **A**lgebra **S**ubprograms): very efficiently preforming common linear algebra problem.

ATLAS (**A**utomatically **T**uned **L**inear **A**lgebra **S**oftware): BLAS implementation with higher performance.

LAPACK (**L**inear **A**lgebra **PACK**age): stands on EISPAC and LINPACK and heavily on BLAS (all written in Fortran) to make them run efficiently on shared-memory vector and parallel processors.

Matlab : is a commercial SW:

- late 1970s, written to access to EISPACK and LINPACK without learning Fortran.
- Then was written in C.
- Then, in 2000, rewritten to use LAPACK.

Mathematica : Commercial SW for symbolic (and numeric of course) mathematical computations.

R : free software environment for statistical computing and graphics.

Python : is a widely used high-level, general-purpose, interpreted, dynamic programming language

Sage : SageMath is a free open-source mathematics software system licensed under the GPL.

- It builds on top of many existing open-source packages: NumPy, SciPy, matplotlib, Sympy, Maxima, GAP, FLINT, R and many more.
- Access their combined power through a common, Python-based language or directly via interfaces or wrappers.
- Mission: Creating a viable free open source alternative to Magma, Maple, Mathematica and Matlab.
- **Examples and Sage cheat sheet:**

2.7.2 Issues on Complexity*

To measure algorithm complexity we need to define a **step**; We adopt the definition of **FLOP** (Floating Point Operation) from the great and very mature reference for matrix computations (Golub and Van Loan, 1996, Sec. 1.2.4): $\square \times \square \quad + \quad \square$ **(almost the inner loop)**

Example 61 (LU factorization) *:*

***LU factorization steps* =**

$$
\begin{aligned}
&= (n)(n-1) + (n-1)(n-2) + \cdots + 2\cdot 1\\
&= \sum_{i=1}^{n} (n-i+1)(n-i)\\
&= \sum_i \Big(i^2 - (2n+1)i + n(n+1)\Big)\\
&= \big(\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n\big) - (2n+1)\big(\frac{1}{2}n(n+1)\big) + n^2(n+1)\\
&= \frac{1}{3}n^3 - \frac{1}{3}n = O(n^3).
\end{aligned}
$$

steps of side of b **=**

$$
\begin{aligned}
&= (n-1) + (n-2) + \cdots + 1\\
&= \frac{1}{2}(n-1)n = \frac{1}{2}n^2 - \frac{1}{2}n = O(n^2).
\end{aligned}
$$

```
sage: var('i,j,k,n');
sage: sum((n-i+1)*(n-i),i,1,n)
1/3*n^3 - 1/3*n
```

Example 62 (Elaboration on Lemma 57 and looping over LT (or UT)) .

Multiplication $A_{m\times n}B_{n\times p}$, $C_{ij} = \sum_k A_{ik}B_{kj}$, mnp *(or* n^3*) steps:*

```
C = 0
for i=1:m
    for j=1:p
        for k=1:n
            C(i,j) = A(i,k)B(k,j) + C(i,j)
```

If both A, B *are LT:* $C_{ij} = 0,\ \forall i < j$, $C_{ij} = \sum_{k=j}^{i} A_{ik}B_{kj},\ \forall j \le i$

```
C = 0
for i=1:n
    for j=1:i // (to access the UT j=i:n)
        for k=j:i //B=0 for k<j, A=0 for i<k
            C(i,j) = A(i,k)B(k,j) + C(i,j)
```

no. of steps =

$$
\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{i}\sum_{k=j}^{i} 1 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{i}(i+1-j) \\
&= \sum_{i=1}^{n}\Big((i+1)i - \frac{1}{2}i(i+1)\Big) = \frac{1}{2}\sum_{i=1}^{n}(i+i^2) \\
&= \frac{1}{2}\Big(\frac{1}{2}n(n+1) + \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n\Big) \\
&= \frac{1}{6}n^3 + \frac{1}{2}n^2 + \frac{1}{3}n.
\end{aligned}
$$

```
sage: var('i,j,k,n');
sage: sum(sum(sum(1, k, j, i), j, 1, i), i, 1, n)
1/6*n^3 + 1/2*n^2 + 1/3*n
```

Example 63 (Matrix round off error and LU partial permutation) *(Golub and Van Loan, 1996, Sec. 3.3). Suppose the PC has a floating point arithmetic with $t = 3$ digits; what is the LU factorization/solution to:*

$$\begin{pmatrix} .001 & 1.00 \\ 1.00 & 2.00 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.00 \\ 3.00 \end{pmatrix}$$

Infinite precision solution (exact):

$$L = \begin{pmatrix} 1 & 0 \\ 1000 & 1 \end{pmatrix}, \; U = \begin{pmatrix} .001 & 1 \\ 0 & -998 \end{pmatrix}, \; LU = \begin{pmatrix} .001 & 1.00 \\ 1 & 2.00 \end{pmatrix} = A, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 500/499 \\ 997/998 \end{pmatrix} = \begin{pmatrix} 1.002004 \\ 0.998998 \end{pmatrix}$$

3-digit precision:

$$L = \begin{pmatrix} 1.00 & 0 \\ 1000 & 1.00 \end{pmatrix}, \; U = \begin{pmatrix} .001 & 1.00 \\ 0 & -1000 \end{pmatrix}, \; LU = \begin{pmatrix} .001 & 1.00 \\ 1.00 & 0.00 \end{pmatrix} = A, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.00 \\ 1.00 \end{pmatrix}$$

some calculation steps:

$$-1000 \times 1 + 2 = -1.00 \times 10^3 + 0.002 \times 10^3 = (-1.00 + 0.00) \times 10^{-3} = -1000.$$

$$1 \times c_1 = 1 \to c_1 = 1; \quad 1000c_1 + c_2 = 3 \to c_2 = -1000; \quad -1000x_2 = -1000 \to x_2 = 1; \quad .001x_1 + x_2 = 1 \to x_1 = 0.$$

3-digit precision with partial pivoting:

$$L = \begin{pmatrix} 1.00 & 0 \\ .001 & 1.00 \end{pmatrix}, \; U = \begin{pmatrix} 1.00 & 2.00 \\ 0 & 1.00 \end{pmatrix}, \; LU = \begin{pmatrix} 1 & 2.00 \\ .001 & 1.00 \end{pmatrix} = A, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.00 \\ .996 \end{pmatrix}$$

Chapter 3

Vector Spaces and Subspaces

3.1 Spaces of Vectors

Definition 64 (A Real Vector Space) *is a set $\mathcal{V}$ of vectors (each is $n-$tuple) over $\mathcal{R}$ with an addition and scalar multiplication on $\mathcal{V}$ such that:*

commutativity $u+v=v+u\in\mathcal{V} \quad \forall u,v\in\mathcal{V}$.

associativity $(u+v)+w=u+(v+w)\in\mathcal{V}$ *and* $(ab)v=a(bv)\in\mathcal{V} \quad \forall u,v,w\in\mathcal{V},\ a,b\in\mathcal{R}$.

additive identity $\exists\,\mathbf{0}\in\mathcal{V}$ *such that* $v+\mathbf{0}=v, \quad \forall v\in\mathcal{V}$.

additive inverse $\forall v\in\mathcal{V}\,\exists w\in\mathcal{V}$ *such that* $v+w=\mathbf{0}$. *(we may denote w by $-v$)*

multiplicative identity $1v=v \quad \forall v\in\mathcal{V}$.

distributive properties $a(u+v)=au+av\in\mathcal{V}$ *and* $(a+b)u=au+bu\in\mathcal{V} \quad \forall u,v\in\mathcal{V},a,b\in\mathcal{R}$.

Hint:

- *Informally: it is the set of vectors which all additions and scalars lay in the set as well.*
- *any linear combination lie in the subspace (from first and last identity)* ■

Example 65 ($\mathcal{R}^2=\{(x_1,x_2)|x_1,x_2\in\mathcal{R}\}$ vs. $\mathcal{V}=\{(x_1,x_2)|-a\le x_1,x_2\le a\}$ is NOT an example) .

Hint:

- We expanded from the visual $n = 3$ to general n.
- We can expand from the set R to any $\mathcal{F}$; the vector space will be defined then over this $\mathcal{F}$. ■
- $x = (x_1, x_2) \in \mathcal{R}^2$ is a point, vector, 2-tuple, element in $\mathcal{R}^2$.
- We can generalize for $\mathcal{R}^n$, or even $\mathcal{C}^n$, or polynomial, or others.
- The human brain cannot visualize or provide geometric models of $\mathcal{R}^n, \ n \geq 4$.
- Edwin A. Abbott, 1884, "Flatland: a romance of many dimensions": can help creatures living in three-dimensional space, such as ourselves, imagine a physical space of four or more dimensions.
- However, we can do mathematics defined $\forall n$ which complies with geometry of $1 \leq n \leq 3$.

Example 66 (Many other spaces of) :

Real: $p = (1, 4, \sqrt(3), -1, 0) \in \mathcal{R}^5$

Complex: $p = (1 + i, -2i, -\sqrt(2) + 3i) \in \mathcal{C}^3$.

Polynomial: $p(z) = a_o + a_1 z + \cdots + a_m z^m$.

3.1.1 Properties of Vector Spaces (seems trivial for R but deep for others!)*

Proposition 67 *For ANY vector space satisfying definition 64 we have the following properties:*

1. *the additive identity is unique.*
2. *the additive inverse of every element is unique.*
3. $0v = \mathbf{0} \quad \forall v \in \mathcal{V}$.
4. $a\mathbf{0} = \mathbf{0} \quad \forall a \in F$.
5. $-1v \; \forall v \in \mathcal{V}$ *is the additive inverse of* v, $(-v)$.

The proof is very trivial:
Proof.

$$\mathbf{0}' = \underleftrightarrow{\mathbf{0}' + \mathbf{0}} = \mathbf{0}$$

$$w = w + \mathbf{0} = \underleftrightarrow{w + (v + w') = (w + v) + w'} = \mathbf{0} + w' = w'.$$

$$0v = (0+0)v = 0v + 0v \longrightarrow 0v = \mathbf{0}$$

$$a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0} \longrightarrow a\mathbf{0} = \mathbf{0}$$

$$v + (-1)v = (1)v + (-1)v = (1-1)v = 0v = \mathbf{0} \longrightarrow (-1)v \text{ is the additive inverse of } v$$

■

3.1.2 Subspaces

Definition 68 *A subset $\mathcal{U}$ of $\mathcal{V}$ is called a subspace of $\mathcal{V}$ if $\mathcal{U}$ is also a vector space (of course using the same addition and scalar as $\mathcal{V}$).*

Example 69 $\mathcal{U} = \left\{(x_1, x_2, 0) | x_1, x_2 \in \mathcal{R}\right\}$ *is a subspace of $\mathcal{R}^3$ since it satisfies all the properties of a space.* ■

Proposition 70 *For any space $\mathcal{V}$ and a subset $\mathcal{U} \subset \mathcal{V}$, $\mathcal{U}$ is a space (or a subspace) if the following hold:*

additive identity $\mathbf{0} \in \mathcal{U}$.

closed under addition $\forall u, v \in \mathcal{U},\ u + v \in \mathcal{U}$.

closed under scalar multiplication $\forall a \in \mathcal{R},\ au \in \mathcal{U}$.

Proof. The proof is obvious since other properties are satisfied immediately on the subset as long as they are satisfied on the whole set. ■

Corollary 71 *The smallest subspace over $\mathcal{R}^n$ is* $\mathbf{0}$.

Example 72 *Which of the following is a subspace (draw):*

- $\mathcal{U} = \left\{(x_1, x_2) | x_1 \in \mathcal{R}, x_2 = ax_1 + b,\ a \neq 0\right\}$, *compare it to* $\mathcal{R}^2$, *then set a condition to be a subspace of* $\mathcal{R}^2$.
  1. $(0,0) \in \mathcal{U} \longrightarrow (0, a \times 0 + b) \in \mathcal{U} \longrightarrow b = 0.$
  2. $(x_1, ax_1) + (x_2, ax_2) = ((x_1 + x_2), a(x_1 + x_2)) \in \mathcal{U}.$
  3. $k(x_1, ax_1) = ((kx_1), a(kx_1))$

- $\mathcal{U} = \left\{(x_1, x_2) | 0 \leq x_1, x_2\right\}.$

- $\mathcal{U} = \left\{(x_1, x_2) | x_1 \in \mathcal{R}, x_2 = ax_1^2,\ a \neq 0\right\}$
  1. $(0,0) \in \mathcal{U}.$
  2. $(x_1, ax_1^2) + (x_2, ax_2^2) = \big((x_1 + x_2), a(x_1^2 + x_2^2)\big) \neq \big((x_1 + x_2), a(x_1 + x_2)^2\big)$

3.1.3 The column space of the matrix A

Definition 73 (Column Space) *of a matrix $A_{m\times n}$, denoted by $\mathcal{C}(A)$, is the vector subspace of $\mathcal{R}^m$ (or probably the whole $\mathcal{R}^m$) consisting of all linear combinations of the matrix columns; i.e., Ax. Said differently:*

$$\mathcal{C}(A) = \left\{Ax \mid \forall x \in \mathcal{R}^n\right\}.$$

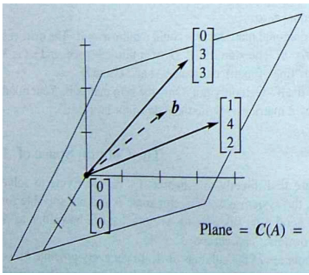
$\mathcal{C}(A)$ ***is the span of the columns of*** A

proof of $\mathcal{C}(A)$ is really a subspace. : $\mathcal{C}(A)$ is really a subspace since $\mathbf{0} \in \mathcal{C}(A)$ by choosing $x = 0$; $Ax_1 + Ax_2 = A(x_1 + x_2) \in \mathcal{C}(A)$; and $a(Ax_1) = A(ax_1) \in \mathcal{C}(A)$. ■

Remark 2 *This recalls the solution of $Ax = b$ exists? b must be in the column space of A.*

Example 74 *What is the column space of the matrix $A = \begin{pmatrix} 1 & 0 \\ 4 & 3 \\ 2 & 3 \end{pmatrix}$?*

It is the set $\mathcal{C}(A) = Ax = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} x_1 + \begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} x_2 \quad \forall x_1, x_2$, which is actually a plane passing through zero.



Example 75 *Describe the column spaces of each of the following:*

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{pmatrix}.$$

It is obvious that all are subspaces of $\mathcal{R}^2$ (probably $\mathcal{R}^2$ itself).

3.2 The Nullspace of A: Solving Ax = 0 and Rx = 0

It is natural to define the row space of a matrix analogously to the column space; **but nothing new!**.

Definition 76 (Row Space $\mathcal{R}(A) \subseteq \mathcal{R}^n$)

$$\mathcal{R}(A) = \left\{x'A \mid \forall x \in \mathcal{R}^m\right\} = \left\{(A'x)' \mid \forall x \in \mathcal{R}^m\right\}$$
$$\mathcal{R}(A) = \mathcal{C}(A'),$$

with no distinction between x and x' (both are in $\mathcal{R}^m$).

Now: it is natural to define a space from ONLY x (not Ax or $x'A$), under some constraint.

Definition 77 (Null Space $\mathcal{N}(A) \subseteq \mathcal{R}^n$) *, and is constructed such that $\mathcal{N}(A) \perp \mathcal{R}(A)$.*

$$\mathcal{N}(A) = \left\{x \mid Ax = \mathbf{0},\ x \in \mathcal{R}^n\right\}.$$

Proof of $\mathcal{N}(A)$ is really a subspace.

$$\begin{aligned} x = \mathbf{0} \quad &\longrightarrow A\mathbf{0} = \mathbf{0} \\ x_1, x_2 \in \mathcal{N}(A) \quad &\longrightarrow \quad Ax_1 = Ax_2 = \mathbf{0} = Ax_1 + Ax_2 = A(x_1 + x_2) \\ x_1 \in \mathcal{N}(A) \quad &\longrightarrow \quad Ax_1 = \mathbf{0} = aAx_1 = A(ax_1). \end{aligned}$$

■

Remark 3 .

1. *It is impossible to have a subspace of $\left\{x \mid Ax = b,\ x \in \mathcal{R}^n\right\}$ except for $b = 0$; why?*

2. *$Ax = 0$ means both: $x \in \mathcal{R}(A)$ and x is a zero linear combination in $\mathcal{C}(A)$.*

Example 78 *What is the null space of* $A = \begin{pmatrix}1 & 2\\ 3 & 6\end{pmatrix}$. *It is of course the solution to* $Ax = 0$ *(by def.):*

$$\begin{pmatrix}1 & 2\\ 3 & 6\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 2\\ 0 & 0\end{pmatrix} \equiv \begin{pmatrix}1 & 2\\ 0 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \mathbf{0} \equiv 1x_1 + 2x_2 = 0 \longrightarrow x = x_2\begin{pmatrix}-2\\ 1\end{pmatrix}, \forall x_2 \in \mathcal{R}. \qquad (E_{21}(-3))$$

$x \perp (1,\, 2)'$ *ONLY. The null space of* A *is the set of vectors constituting the line.*

Example 79 *What is the null space of the matrix:* $x_1 + 2x_2 + 3x_3 = 0$. *Here:* $A = (1, 2, 3)$. *No pivot cancellation:*

$$x = \begin{pmatrix}-2x_2 - 3x_3\\ x_2\\ x_3\end{pmatrix} = x_2\begin{pmatrix}-2\\ 1\\ 0\end{pmatrix} + x_3\begin{pmatrix}-3\\ 0\\ 1\end{pmatrix}$$

The solution is the set of all linear combinations of these two $(2 = 3 - 1)$ *simple vectors;* ***A PLANE:*** *let's draw it.*

Example 80 *Suppose* $A =$

$$A = \begin{pmatrix}1 & 2 & 3\\ 1 & 1 & -4\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 2 & 3\\ 0 & -1 & -7\end{pmatrix} \longrightarrow x_2 = -7x_3, \quad x_1 = 11x_3 \longrightarrow x = x_3\begin{pmatrix}11\\ -7\\ 1\end{pmatrix}.$$

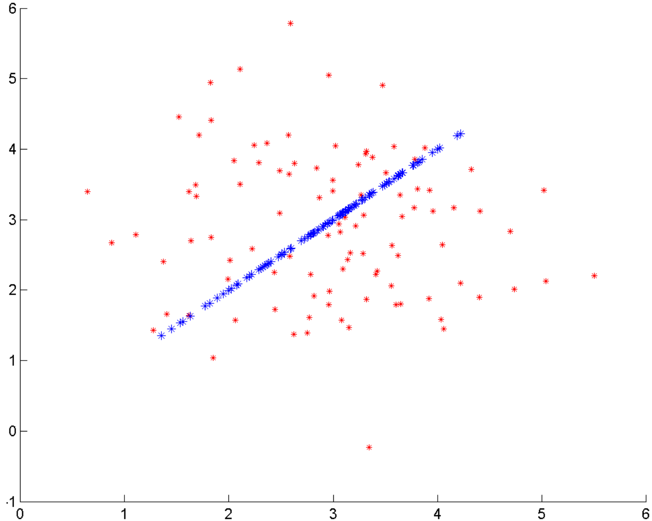
Much easier to continue from U *to the reduced echelon form* R*:*

$$\begin{pmatrix}1 & 2 & 3\\ 0 & -1 & -7\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 0 & -11\\ 0 & -1 & -7\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 0 & -11\\ 0 & 1 & 7\end{pmatrix}$$

So, the solution is the set of all linear combination of this single $(1 = 3 - 2)$ *vector;* ***A Line:*** *let's see Sage.*

Example 81 (Motivation from data science) :

- *Data reduction and compression.*
- *Data Interpretation.*
- *Data modeling and prediction.*



3.2.1 Systematic solution using pivot columns, free columns, and reduced echelon form

Example 82 *Find $\mathcal{N}(A)$:*

$$\begin{pmatrix}1 & 2 & 2 & 4\\ 3 & 8 & 6 & 22\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 2 & 2 & 4\\ 0 & 2 & 0 & 10\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 0 & 2 & -6\\ 0 & 2 & 0 & 10\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 0 & 2 & -6\\ 0 & 1 & 0 & 5\end{pmatrix}.$$

In reduced echelon form, we get r pivot variables p and $n-r$ free variables f; in the form of $p = -\sum \alpha f$:

$$x_1 = -2x_3 + 6x_4$$
$$x_2 = 0x_3 - 5x_4$$
$$x = \begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} = x_3\begin{pmatrix}-2\\0\\1\\0\end{pmatrix} + x_4\begin{pmatrix}6\\-5\\0\\1\end{pmatrix}$$

Example 83 *After pivot cancellation of A:*

$$\begin{pmatrix}1 & 0 & 0 & a & c\\ 0 & 1 & 0 & b & d\\ 0 & 0 & 1 & 0 & e\\ 0 & 0 & 0 & 0 & 0\end{pmatrix} \longrightarrow \quad x = x_4\begin{pmatrix}-a\\-b\\0\\1\\0\end{pmatrix} + x_5\begin{pmatrix}-c\\-d\\-e\\0\\1\end{pmatrix} \qquad \begin{pmatrix}1 & 0 & a & 0 & c\\ 0 & 1 & b & 0 & d\\ 0 & 0 & 0 & 1 & e\\ 0 & 0 & 0 & 0 & 0\end{pmatrix} \longrightarrow \quad x = x_3\begin{pmatrix}-a\\-b\\1\\0\\0\end{pmatrix} + x_5\begin{pmatrix}-c\\-d\\0\\-e\\1\end{pmatrix}$$

3.2.2 Gauss Elimination Algorithm: revisited and detailed

```
i = 0; j = 0; //previous pivot location
while( (i<m) && (j<n) ){
    i++; j++;
    do{
        I = argmax( |A(k,j)|, i<=k<=m);
        j+= !A(I,j);
    }while( !A(I,j) && (j<=n) );

    if(A(I,j)){//pivot or reached boundary
        Swap(R_i, R_I);
        PivotElimination(i,j);
    }
}
```

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Corollary 84 (Gauss elimination algorithm) .
For $A_{m\times n}$ that produces r pivots:

1. $R_{ij} = 0,\ \forall i > I,\ j < J,\ R(I,J)$ *is a pivot.*
2. *the number of pivots r, number of column pivots, and number of row pivots are all equal.*
3. $r \le m,n \quad \equiv \quad r \le min(m,n)$.
4. *the $m-r$ non-pivot rows are all zeros and are deferred to the end of R.*
5. *the $n-r$ non-pivot columns have zeros under the previous pivot.*

Proof. It is trivial and is already a bi-product from the construction of elimination!

■

Lemma 85 *In pivot cancellation, a column will have no pivots if and only if it is a linear combination from preceding columns. A row will be zero if and only if it is a linear combination of preceding rows.*

Proof.
For columns:

$$\begin{pmatrix} 1 & \mathbf{0} \\ -A_1/a_{11} & \mathbf{I}. \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & \alpha a_{11}+\beta a_{21} & \cdots \\ A_1 & A_2 & \cdots & \alpha A_1 + \beta A_2 & \cdots \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \cdots & \alpha a_{11}+\beta a_{21} & \cdots \\ \mathbf{0} & A_2 - A_1 a_{21}/a_{11} & \cdots & \beta(A_2 - A_1 a_{21}/a_{11}) & \cdots \end{pmatrix}$$

second pivot cancellation will not provide pivots in the linear combination column.

For rows:

$$\begin{pmatrix} a_{11} & R_1 \\ a_{21} & R_2 \\ \vdots & \vdots \\ \alpha a_{11}+\beta a_{21} & \alpha R_1 + \beta R_2 \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11} & R_1 \\ 0 & R_2 - R_1(a_{21}/a_{11}) \\ \vdots & \vdots \\ 0 & (\alpha R_1 + \beta R_2) - R_1\big(\alpha a_{11} + \beta a_{21}\big)/a_{11} \end{pmatrix} = \begin{pmatrix} a_{11} & R_1 \\ 0 & R_2 - R_1(a_{21}/a_{11}) \\ \vdots & \vdots \\ 0 & \beta\big(R_2 - R_1(a_{21}/a_{11})\big) \end{pmatrix}$$

■

Definition 86 (Rank of a matrix) *is defined as the number of its pivots r.*
Later, an equivalent definition is provided and we will show that r is the number of independent columns, independent rows, etc. ■

Lemma 87 *After Gauss elimination of A to produce the echelon (or reduced echelon) form R:*

1. *All pivot columns of R are linearly independent; their corresponding columns of A are linearly independent as well.*
2. *All non-pivot columns of R are linear combination of preceding columns; same apply for matrix A.*
3. *All pivot rows of R are linearly independent. their corresponding rows of A are linearly independent as well.*
4. *All non-pivot rows of R (the zero rows) are linear combination from the pivot rows; same apply for matrix A*

Proof. We arrange

1. We assume that $\exists\ \alpha$, a linear combination of pivot columns, such that $R\alpha = \mathbf{0}$

$$R\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_r\\ 0\\ 0\\ \vdots\\ 0\end{pmatrix} = \mathbf{0} \quad \to \alpha \in \mathcal{N}(R) \quad \to \alpha = x_{r+1}\begin{pmatrix}*\\ \vdots\\ *\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix} + x_{r+2}\begin{pmatrix}*\\ \vdots\\ *\\ 0\\ 1\\ \vdots\\ 0\end{pmatrix} + \cdots + x_n\begin{pmatrix}*\\ \vdots\\ *\\ 0\\ 0\\ \vdots\\ 1\end{pmatrix} \quad \to x_i = 0,\ r+1 \le i \le n.$$

which means $\alpha = \mathbf{0}$; a contradiction.

2. Since we have just proven that $\alpha \notin \mathcal{N}(A)$; then $A\alpha \neq \mathbf{0}$. Therefore, the corresponding columns of A are linearly independent as well.

∎

Remark 4 *:*

- $\mathcal{N}(A) = \mathcal{N}(U) = \mathcal{N}(R)$ *of course, since pivot cancellation will not change the* $\mathbf{0}$ *vector at the R.H.S.*

- $\mathcal{C}(A) \neq \mathcal{C}(U) \neq \mathcal{C}(R)$*; simply:*

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \quad \mathcal{C}(A) = a\begin{pmatrix} 1 \\ 2. \end{pmatrix} \qquad ,R = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix},\ \mathcal{C}(R) = a\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

  We will come later to how to find exactly $\mathcal{C}(A)$ *and* $\mathcal{R}(\mathcal{A})$.

- *The number of vectors in* $\mathcal{N}(A)$ *is itself the number of linear combinations of columns of* A *that gives* $\mathbf{0}$.

3.3 3.3 The Complete Solution to Ax = b

3.4 3.4 Independence, Basis and Dimension

3.5 3.5 Dimensions of the Four Subspaces

Bibliography

Golub, G. H., Van Loan, C. F., 1996. Matrix computations, 3rd Edition. Johns Hopkins University Press, Baltimore.

Schott, J. R., 2005. Matrix analysis for statistics, 2nd Edition. Wiley, Hoboken, N.J.

Searle, S. R., 1982. Matrix algebra useful for statistics. Wiley, New York.

Strang, G., 2016. Introduction to linear algebra, 5th Edition. Wellesley-Cambridge Press.