Linear algebra: theory and practice and programming

Dr. Yihsiang Liow (October 21, 2025)

Contents

debug: intro.tex

Chapter 1

Introduction

1.1 Goal

TO READER: This set of notes on linear algebra is not meant for beginners (yet). The goal is to review the main facts about linear algebra. However a beginner who has not studied linear algebra (from a linear algebra class or from some of my classes – 380 computer graphics) can benefit if he/she studies this set of notes with a beginner's linear algebra textbook. Some of the facts are beyond the usual undergraduate linear algebra course. Think of this set of notes as a summary. Even for an undergraduate who has experience with linear algebra, you can/should study this set of notes together with a linear algebra textbook as a cross reference. This set of notes is also different from other undergraduate linear algebra textbooks because I'm including software computation of linear algebra with the use of linear algebra libraries for C/++ and python.

Te second goal is to review linear algebra libraries for C/C++ and python.

1. Python: numpy and scipy
2. Python: OpenCV [see opencv notes]
3. C++: BLAS
4. C++: Boost uBLAS
5. C++: Eigen3
6. C++: glm [see ciss380 notes]
7. C++: OpenCV [see opencv notes]

and include information how to use the libraries. I will also use

1. sagemath for symbolic computation (including the linear algebra library)

debug: summary-linear-algebra.tex

Chapter 2

Summary of linear algebra

NOTATION: As with my other notes (350, 358, etc.), most textbooks/papers (as usual) will index starting with 1, but I'll start indexing with 0. So an array of size n in most textbooks would look like `x[1..n]` whereas for me I'm write `x[0..n - 1]`. In regular textbook, an n–dimensional row vector would look like $[v_1, ..., v_n]$ whereas for my notes it would look like $[v_0, ..., v_{n-1}]$. There's hardly any difference since the translation back and forth is immediate. I'm indexing from 0 because it's closer to programming.

2.1 Vector spaces

Let $R = (R, +_R, \cdot_R, 0_R, 1_R)$ be a ring. $M = (M, +_M, \cdot_M, 0_M)$ is an R–module if $(M, +_M, 0_M)$ is an abelian group and $\cdot_M : R \times M \to M$ satisfies

$$\begin{aligned} r \cdot_M (m +_M m') &= r \cdot_M m +_M r \cdot_M m' \\ (r +_R r') \cdot_M m &= r \cdot_M m +_M r' \cdot_M m \\ (r \cdot_R r') \cdot_M m &= r \cdot_M (r' \cdot_M m) \\ 1_R \cdot_M m &= m \end{aligned}$$

for all $r, r' \in R$ and $m, m' \in M$. $\cdot_M$ is called the scalar multiplication of M. If it's clear which multiplication or addition is applicable, removing the subscripts gives us:

$$\begin{aligned} r \cdot (m + m') &= r \cdot m + r \cdot m' \\ (r + r') \cdot m &= r \cdot m + r' \cdot m \\ (r \cdot r') \cdot m &= r \cdot (r' \cdot m) \\ 1 \cdot m &= m \end{aligned}$$

A value in R is called a **scalar**. If $r \in R$ and $m \in M$ is an R–module, the multiplication $r\dot{m}$ is called a **scalar multiplication**.

scalar

scalar multiplication

Note that there is no mention of multiplication in M in the definition above. (If M has its own multiplication, together with some extra axioms, M is an R–algebra.)

The definition of an R–module does not require R to be a commutative ring. Therefore one can define a similar concept of modules over non–commutative rings, over commutative rings, over division rings (rings where every non-zero element is invertible), and over fields (commutative rings where every non-zero elementis invertible).

In particular, if R is a field K, then the above K–module is also called a K–**vector space**. Instead of saying V is a K-vector space, it's also common to say V is a "vector space **over** K".

vector space

over

Example 2.1.1.

(a) Every ring R is an R–module. Therefore $\mathbb{Z}$ is a $\mathbb{Z}$–module, $\mathbb{Q}$ is a $\mathbb{Q}$–module, $\mathbb{R}$ is an $\mathbb{R}$–module, and $\mathbb{C}$ is a $\mathbb{C}$–module, etc.
(b) If I is an ideal of ring R, then R/I is an R–module. Let $N > 0$. Then $\mathbb{Z}/N$ is a $\mathbb{Z}$–module.
(c) Every abelian group is a $\mathbb{Z}$–module.
(d) If R'/R is a ring extension (i.e., R is a subring of R'), then R' is an R–module. $\mathbb{Q}$ is a $\mathbb{Z}$–module, $\mathbb{R}$ is a $\mathbb{Q}$– and $\mathbb{Z}$–module. $\mathbb{C}$ is an $\mathbb{R}$–, $\mathbb{Q}$– and $\mathbb{Z}$–module.
(e) If L/K is a field extension (i.e., K is a subfield of L), then L is a K–vector space. Repeating part of the above, $\mathbb{R}$ is a $\mathbb{Q}$–vector space and $\mathbb{C}$ is an $\mathbb{R}$–vector space.
(f) If R is a ring, $R[x]$ (the polynomial ring over R), $R[[x]]$ (the power series ring over R) are R–modules. Likewise for Laurent power series with finitely many negative powers and Laurent power series with infinitely many negative powers. If R is a field, the above modules are vector spaces.

Example 2.1.2.

(a) Let X be a set and S be a ring. The set of functions $X \to S$ is an S–module where if $f, g : X \to S$, then $f + g : X \to S$ is defined (obviously) as

$$(f + g)(x) = f(x) + g(x)$$

and if $s \in S$, then $s \cdot f : X \to S$ is defined as

$$(s \cdot f)(x) = s \cdot (f(x))$$

(b) The set of functions $\mathbb{N} \to \mathbb{R}$ satisfying a homogeneous linear recurrence relation of fixed degree d is a $\mathbb{Z}$–module and an $\mathbb{R}$–vector space.
(c) The set of $\mathbb{R} \to \mathbb{R}$ functions which are d times differentiable $C^d(\mathbb{R})$ and satisfy a homogeneous linear differential equation of order d is a $\mathbb{Z}$–module and an $\mathbb{R}$–vector space.

Example 2.1.3. If R is a ring, then R^n is an R–module where the addition in R^n is defined coordinate-wise

$$(r_0, ..., r_{n-1}) + (r'_0, ..., r'_{n-1}) = (r_0 + r'_0, ..., r_{n-1} + r'_{n-1})$$

(where $+$ on the right is the ring addition of R) and the scalar multiplication is defined as

$$r \cdot (r_0, ..., r_{n-1}) = (r \cdot r_0, ..., r \cdot r_{n-1})$$

(where $\cdot$ on the right is the ring multiplication of R) $\mathbb{Z}^n$ is a $\mathbb{Z}$–module, $\mathbb{R}^n$ is an $\mathbb{R}$–vector space. $\mathbb{C}^n$ is a $\mathbb{Z}$–module and $\mathbb{R}$– and $\mathbb{C}$–vector space.

I will call $\mathbb{R}^n$ the **standard vector space** over $\mathbb{R}$ of dimension n. (See later for dimension.) Elements of $\mathbb{R}^n$ will be written as row vectors $[v_0, ..., v_{n-1}]$ or as column vectors $\begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}$. standard vector space

Example 2.1.4. $\mathrm{Mat}_{r,c}(R)$ is also an R–module. The addition of matrices $A = [a_{i,j}]_{0 \le i < r, 0 \le j < c}$ and $B = [b_{i,j}]_{0 \le i < r, 0 \le j < c}$ is $C = [c_{i,j}]_{0 \le i < r, 0 \le j < c}$ where

$$c_{i,j} = a_{i,j} + b_{i,j}$$

and scalar multiplication of $c \in R$ and A, $C = cA$ is defined by $C = [c_{i,j}]_{0 \le i < r, 0 \le j < c}$ where

$$c_{i,j} = c \cdot a_{i,j}$$

If $r = c$, $\mathrm{Mat}_{r,c}(R)$ is abbreviated to $\mathrm{Mat}_r(R)$.

$0_{r,c}$ denotes the zero matrix with size $r \times c$. If $r = c$, $0_{r,c}$ is abbreviated to 0_r.

If $A = [a_{i,j}]_{0 \le i < r,\ 0 \le j < c}$, then $-A$ is the $r \times c$ matrix where the (i, j)–entry is $-a_{i,j}$. As usual, if A and B are $r \times c$ matrices, $A - B$ is defined to be $A + (-B)$.

With the above, $\text{Mat}_{r,c}(R)$ is an R–module since $\text{Mat}_{r,c}(R)$ is the same as $R^{r\times c}$. $\text{Mat}_{r,c}(\mathbb{R})$ is an $\mathbb{R}$–vector space. $\text{Mat}_{r,c}(\mathbb{C})$ is an $\mathbb{R}$– and $\mathbb{C}$–vector space.

The product of matrices $A = [a_{i,j}]_{0\leq i<p,0\leq j<q}$ and $B = [b_{i,j}]_{0\leq i<q,0\leq j<r}$ is $C = [c_{i,j}]_{0\leq i<p,\ 0\leq j<r}$ where

$$c_{i,j} = \sum_{k=0}^{q-1} a_{i,k}b_{k,j}$$

$I_{r,c}$ is the $r \times c$ matrix with 1s on the diagonal and 0s off the diagonal, i.e., $I_{r,c} = [a_{i,j}]_{0\leq i<r,0\leq j<c}$ where

$$a_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Now restrict $\text{Mat}_{r,c}(R)$ to square matrices.

If A is $n \times n$, then the inverse of A denoted by A^{-1}, if it exists, is an $n \times n$ matrix such that

$$AA^{-1} = I_n = A^{-1}A$$

A is said to be invertible if A^{-1} exists. A matrix is **singular** if it is not invertible. Otherwise the matrix is **nonsingular**. With the above $\text{Mat}_n(R)$ is a noncommutative ring where the additive identity is 0_n and the multiplicative identity is I_n. This means that $\text{Mat}_n(R)$ is both a ring and an R–module and is in fact an R–algebra. The set of invertible matrices of $\text{Mat}_n(R)$ is denoted by $\text{GL}_n(R)$, the **general linear group** of size n with values in R. $\text{GL}_n(R)$ is the unit group of $\text{Mat}_n(R)$.

singular

nonsingular

general linear group

If A and B are matrices of sizes $p\times q$ and $r\times s$, then $A+B$ is defined and has size $p \times q$ only when $p = r, q = s$ and AB is defined and has size $p \times r$ only when $q = r$. Otherwise $A + B$ and AB are not defined.

if $K = \mathbb{C}$, $\overline{A}$ is the complex conjugate of A where the (i, j)–entry if $\overline{a_{i,j}}$, i.e., the complex conjugate of $a_{i,j}$.

Proposition 2.1.1. *Let A, B, C be matrices. Each of the following assumes the sizes are compatible.*

(a) $(A + B) + C = A + (B + C)$.
(b) $A + 0 = A = 0 + A$
(c) $A + (-A) = 0 = A + (-A)$ *where*
(d) $(AB)C = A(BC)$ *if A, B, C has sizes $p \times q, q \times r, r \times s$.*

(e) $AI = A = IA$ *where* A *is a square matrix.*
(f) $A(B + C) = AB + AC$

In particular the set of matrices over a fixed field and with fixed size n, $\mathrm{Mat}_n(K)$, besides being a K–vector space, is also a ring. The scalar multiplication works well with the ring multiplication, i.e., $c(AB) = (cA)B = A(cB)$. We say that $\mathrm{Mat}_n(K)$ is a K–algebra. (Memory aid: algebra = ring + vector space.)

Example 2.1.5.

(a) The set of solutions $(x, y, z) \in \mathbb{R}^3$ satisfying a linear homogeneous system

$$\begin{aligned} 3x + 4y + z &= 0 \\ -x + 2y + 11z &= 0 \end{aligned}$$

i.e.,

$$\begin{bmatrix} 3 & 4 & 1 \\ -1 & 2 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

i.e., is a vector space. In general the solutions $v \in \mathbb{R}^c$ to

$$Av = 0$$

where $A \in \mathrm{Mat}_{r,c}(\mathbb{R})$ forms a vector space in $\mathbb{R}^c$, the kernel of A. In the above $\mathbb{R}$ can be replaced by any field.

(b) For a fixed $n > 0$, the solutions of the $n \times n$ magic square problem with values in a ring K is a K–vector space. (A solution means "all row sums, all column sum, both diagonal sums have the same value") There are n^2 variables and $2n + 2$ linear equations.

Example 2.1.6.

(a) Let $C^0([a, b])$ be the set of continuous functions $[a, b] \to \mathbb{R}$ where $a < b$ are fixed real numbers. Then $C^0([a, b])$ is an $\mathbb{R}$–vector space.

(b) The set X of differentiable functions $[a, b] \to \mathbb{R}$ is also an $\mathbb{R}$–vector space. $C^1([a, b])$ is the set of functions which are differentiable wth continuous derivative. $C^1([a, b])$ is a subset of X and is an $\mathbb{R}$–vector space (in fact it's a subspace – see later.)

From now on, I will mainly focus on vector spaces of finite dimension, although some of the information below applies to general vector spaces.

2.2 Subspaces and quotient spaces

Let V be a K–vector space.

A K–**subspace** (or sub–K–vector space) of V is a subset of V that is a K–vector space using the same operators of V. If V' is a subspace of V, we write

subspace

$$V' \leq V$$

(a) 0 and V are subspaces of V. Here 0 means $\{0\}$.
(b) Reflexitivity: V is a subspace of V.
(c) Transitivity: If V is a subspace of V' and V' is a subspace of V'', then V is a subspace of V''.
(d) If V, V' are subspaces of V'', then $V \cap V'$ is a subspace of V''.

If V', V'' are subspaces of V, then

$$V' + V'' = \{v' + v'' \mid v' \in V', \;\; v'' \in V''\}$$

is also a subspace.

If V' is a subspace of V, then V/V' the **quotient space** is a vector space. Here V/V' as a set is defined using the quotient of V by V' as abelian group, i.e.,

quotient space

$$(v + V') + (v' + V') = (v + v') + V'$$

for $v, v' \in V$ and the scalar multiplication of V/V' is defined as

$$c \cdot (v + V') = (cv) + V'$$

where c is a scalar. The map $V \to V/V'$, $v \mapsto v + V'$ is the **standard quotient map** of $V' \leq V$. This is analogous to $\mathbb{Z} \to Z/N$.

standard quotient map

2.3 Homomorphism and isomorphism

Let V, V' be K–vector spaces. A function $f : V \to V'$ is a **homomorphism** of K–vector spaces if homomorphism

$$f(v + v') = f(v) + f(v')$$
$$f(cv) = cf(v)$$

f is also called a K–**linear map**. linear map

If $f : V \to V'$ is K–linear, then the image of f, denoted by $\mathrm{im}(k)$ is a subspace of V' and the kernel of f,

$$\ker(f) = \{v \in V \mid f(v) = 0\}$$

is a subspace of V. Therefore we have a quotient space of $V/\ker f$.

An **isomorphism** of vector spaces is a bijective homomorphism of vector spaces. If f is an isomorphism of vetor spaces, so is f^{-1}. If $f : V \to V'$ is a homomorphism of vector spaces. isomorphism

If we have have a homomorphism of vector spaces $f : V \to V'$, then we have an exact sequence of vector spaces:

$$0 \to \ker k \to V \to \mathrm{im}\, f \to 0$$

the second map is inclusion (hence 1–1) and the third is the quotient map (hence onto). In general if

$$0 \to U \to V \to W \to 0$$

is an exact sequence of vector spaces (i.e., second map is 1–1 and third map is onto), it splits, i.e.,

$$V \simeq U \oplus W$$

(see section on direct sum $\oplus$.)

Example 2.3.1. If $A \in \mathrm{Mat}_{m,n}(K)$, then $A : \mathbb{R}^n \to \mathbb{R}^m$ is a K–linear map.

Example 2.3.2. Recall that $C^0([a, b])$ is the set of continuous functions $[a, b] \to \mathbb{R}$ and $C^1([a, b])$ is the set of differentiable functions with continuous derivatives on $[a, b] \to \mathbb{R}$ where $a < b$ are fixed real numbers. The **integral operator**

integral operator

$$\int_a^b : C^0(\mathbb{R}) \to \mathbb{R}$$
$$\int_a^b (f) = \int_a^b f(x)\ dx \ \text{(Riemann integral)}$$

is an $\mathbb{R}$–linear map. The **differentiation operator**

differentiation operator

$$\frac{d}{dx} : C^1([a, b]) \to C^0([a, b])$$
$$\frac{d}{dx}(f) = \frac{df}{dx}$$

is an $\mathbb{R}$–linear map.

2.4 Product spaces

$\mathbb{R}^1$ is a vector space. You can create the product (as set) $\mathbb{R}^1 \times \mathbb{R}^1$ with elements (u, v) where $u, v \in \mathbb{R}$. This is also a vector space where addition and scalar multiplication are defined (obviously) as

$$\begin{aligned}(u, v) + (u', v') &= (u + u', v + v')\\ c(u, v) &= (cu, cv)\end{aligned}$$

Clearly $\mathbb{R}^1 \times \mathbb{R}^1$ is the same as $\mathbb{R}^2$ (i.e. they are isomorphic).

You can also create $\mathbb{R}^1 \times \mathbb{R}^2$ out of two pieces $\mathbb{R}^1$ and $\mathbb{R}^2$ in a similar way:

$$\begin{aligned}(u, (v, w)) + (u', (v', w')) &= (u + u', (v, w) + (v', w'))\\ c(u, (v, w))) &= (cu, c(v, u))\end{aligned}$$

This is clearly the same as $\mathbb{R}^3$ (they are isomorphic): every $(u, (v, w))$ of $\mathbb{R}^1 \times \mathbb{R}^2$ corresponds to (u, v, w) of $\mathbb{R}^3$.

For $\mathbb{R}^1 \times \mathbb{R}^2$, which is made up of $\mathbb{R}^1$ and $\mathbb{R}^2$ as components, it contains $\mathbb{R}^1 \times \{(0, 0)\}$ and $\{0\} \times \mathbb{R}^2$. $\mathbb{R}^1 \times \{(0, 0)\}$ is isomorphic to $\mathbb{R}^1$:

$$(u, (0, 0)) \mapsto u$$

$\mathbb{R}^1 \times \mathbb{R}^2$ also contains $\{0\} \times \mathbb{R}^2$ which is isomorphic to $\mathbb{R}^2$:

$$(0, (u, v)) \mapsto (u, v)$$

We sya that $\mathbb{R}^1 \times \mathbb{R}^2$ is an external product/sum of $\mathbb{R}^1$ and $\mathbb{R}^2$; external in the sense that the two pieces $\mathbb{R}^1$ and $\mathbb{R}^2$ are not inside $\mathbb{R}^1 \times \mathbb{R}^2$. We also say that $\mathbb{R}^1 \times \mathbb{R}^2$ is an internal product/sum of $\mathbb{R}^1 \times \{(0, 0)\}$ and $\{0\} \times \mathbb{R}^2$; the notation is $\mathbb{R}^1 \times \mathbb{R}^2 = \mathbb{R}^1 \times \{(0, 0)\} \oplus \{0\} \times \mathbb{R}^2$. Internal in the sense that the two pieces in this case are inside $\mathbb{R}^1 \times \mathbb{R}^2$.

Geometrically and pictorially, inside $\mathbb{R}^1 \times \mathbb{R}^2$, $\mathbb{R}^1 \times \{(0, 0)\}$ are position vectors with end points on the x–axis while $\{0\} \times \mathbb{R}^2$ are position vectors with end points on the yz–plane.

Both views of breaking up $\mathbb{R}^1 \times \mathbb{R}^2$ into components, external and internal, are equally important. But both ways are essentially the same. From the external product components $\mathbb{R}^1$ and $\mathbb{R}^2$, you can get the internal product components $\mathbb{R}^1 \times \{(0, 0)\}$ and $\{0\} \times \mathbb{R}^2$ and both sets of components matches up in the following sense.

The component $\mathbb{R}^1$ is inside $\mathbb{R}^1 \times \mathbb{R}^2$ through it's image $\mathbb{R}^1 \to \mathbb{R}^1 \times \mathbb{R}^2$, $c \mapsto (c, (0,0))$, i.e., $\mathbb{R}^1$ appears inside $\mathbb{R}^1 \times \mathbb{R}^2$ as $\mathbb{R}^1 \times \{(0,0)\}$. The component $\mathbb{R}^2$ is inside $\mathbb{R}^1 \times \mathbb{R}^2$ through it's image $\mathbb{R}^2 \to \mathbb{R}^1 \times \mathbb{R}^2$, $(u,v) \mapsto (0,(u,v))$, i.e., $\mathbb{R}^2$ appears inside $\mathbb{R}^1 \times \mathbb{R}^2$ as $\{0\} \times \mathbb{R}^2$. $\mathbb{R}^1$ appears in $\mathbb{R}^1 \times \mathbb{R}^2$ as an external component while $\mathbb{R}^1 \times \{(0,0)\}$ appears inside $\mathbb{R}^1 \times \mathbb{R}^2$ as an internal component. But they are both the same (i.e., they are isomorphic). Note that the two components inside $\mathbb{R}^1 \times \mathbb{R}^2$ only has $(0,(0,0))$ in common:

$$\mathbb{R}^1 \times \{(0,0)\} \cap \{0\} \times \mathbb{R}^2 = \{(0,(0,0)\}$$

The following is the generalization of the above example.

If V, V' are K–vector spaces then $V \times V'$ is also a K–vector space in the obvious way:

$$(u,u') + (v,v') = (u+u', v+v')$$
$$c(u,v) = (cu, cv)$$

$V \times V'$ is the **external product/sum space** of V and V'. There are obvious K–linear map from V to $V \times V'$ and V' to $V \times V'$: $V \to V \times V'$, $f(u) = (u, 0)$ and $V' \to V \times V'$, $f'(v) = (0, v)$. These two maps, f, f' are injective and their images $\operatorname{im} f, \operatorname{im} f'$ are subspaces of $V \times V'$, $\operatorname{im} f \simeq V$, $\operatorname{im} f' \times V'$, $\operatorname{im} f + \operatorname{im} f' = V \times V'$ and $\operatorname{im} f \cap \operatorname{im} f' = 0$.

external product/sum space

Let U'' be a K–vector space. If U'' contains two vector spaces U, U' such that $U + U' = U''$ and $U \cap U' = 0$, then we write $U'' = U \oplus U'$ and say that U'' is the **internal product/sum space** of U and U'.

internal product/sum space

I'll just call the above two constructions $\times$ and $\oplus$ the **direct product** and **direct sum**.

direct product

direct sum

The following is the generalization of the above example.

If V, V' are K–vector spaces then $V \times V'$ is also a K–vector space in the obvious way:

$$(u,u') + (v,v') = (u+u', v+v')$$
$$c(u,v) = (cu, cv)$$

$V \times V'$ is the **external product/sum space** of V and V'. There are obvious K–linear map from V to $V \times V'$ and V' to $V \times V'$: $V \to V \times V'$, $f(u) = (u, 0)$ and $V' \to V \times V'$, $f'(v) = (0, v)$. These two maps, f, f' are injective and their images $\operatorname{im} f, \operatorname{im} f'$ are subspaces of $V \times V'$, $\operatorname{im} f \simeq V$, $\operatorname{im} f' \times V'$,

external product/sum space

$\operatorname{im} f + \operatorname{im} f' = V \times V'$ and $\operatorname{im} f \cap \operatorname{im} f' = 0$.

Let U'' be a K–vector space. If U'' contains two vector spaces U, U' such that $U + U' = U''$ and $U \cap U' = 0$, then we write $U'' = U \oplus U'$ and say that U'' is the **internal product/sum space** of U and U'.

internal product/sum space

I'll just call the above two constructions $\times$ and $\oplus$ the **direct product** and **direct sum** respectively.

direct product

direct sum

Direct product and direct sum are essentially equivalent in the following sense. If V'' is an external direct product of V and V', then it is an internal sum of some U, U' where $V \simeq U$, $V' \simeq U'$, and $U \cap U' = 0$. If V'' is an internal sum of U, U', then V'' is an external direct product of some V, V' where $V \simeq U$ and $V' \times U'$.

2.5 Linear span

Let X be a set of vectors of V. The (linear) **span** of X, denoted by $\langle X \rangle$, is defined to be

span

$$\langle X \rangle = \left\{ \sum_{i=0}^{n} c_i v_i \,\middle|\, n \geq 0, \text{ and } c_0, ..., c_{n-1} \in K, \text{ and } v_0, ..., v_{n-1} \in X \right\}$$

The expression $\sum_{i=0}^{n} c_i v_i$ is called a **linear combination** of $v_0, ..., v_{n-1}$. Therefore the span of X is the set of all finite linear combinations of X.

linear combination

$\langle X \rangle$ is the smallest subspace of V containing X. $\langle X \rangle$ can also be described as the intersection of all subspaces of V that contains X:

$$\langle X \rangle = \bigcap_{X \subseteq V' \leq V} V'$$

Note that if V', V'' are subspaces of V, then $V' \cup V''$ is not necessarily a subspace of V. $\langle V' \cup V'' \rangle$ is a subspace of V:

$$\begin{aligned} V' \cap V'' \leq V' \leq \langle V' \cup V'' \rangle \\ V' \cap V'' \leq V'' \leq \langle V' \cup V'' \rangle \end{aligned}$$

$V' \cap V''$ is the largest subspace of V' and V'' and $\langle V', V'' \rangle$ is the smallest subspace of V containing V' and V''.

2.6 Linear independence

A finite set of vectors of V, say $v_0, v_1, ..., v_{n-1}$ are **linearly dependent** if there are scalar elements $c_0, ..., c_{n-1}$, not all zeros, such that

linearly dependent

$$\sum_{i=0}^{n-1} c_i v_i = 0$$

$v_0, v_1, ..., v_{n-1}$ are **linearly independent** if

linearly independent

$$\sum_{i=0}^{n-1} c_i v_i = 0 \implies c_0 = \ldots = c_{n-1} = 0$$

where $c_i \in K$.

By the way for the case of infinite sets, the sum is finite, i.e., elements of $X \subseteq V$ are linearly independent if

$$\sum_{i=0}^{n-1} c_i v_i = 0 \implies c_0 = \ldots = c_{n-1} = 0$$

holds for all finite $n > 0$, $c_i \in K$ (for $0 \leq i < n$) and $v_i \in X$ (for $0 \leq i < n$).

Example 2.6.1. In $\mathbb{R}^n$ as $\mathbb{R}$–vector space, $\{e_i \mid 0 \leq i < n\}$ are linearly independent where e_i is made up of 0s except for a 1 at index i. In $\mathrm{Mat}_{r,c}(\mathbb{R})$ as $\mathbb{R}$–vector space, $\{e_{i,j} \mid 0 \leq i < n,\ 0 \leq j < n\}$ are linearly independent where $e_{i,j}$ is made up of 0s except for a 1 at index (i, j).

Example 2.6.2. In $\mathbb{R}$ as $\mathbb{Q}$–vector space, $\{1, \pi, \pi^2\}$ are linearly independent.

Example 2.6.3. In the $\mathbb{R}$–vector space of functions $\mathbb{R} \to \mathbb{R}$,

(a) $\{1, x, \ldots, x^k\}$ (for $k \geq 0$) are linearly independent.
(b) $\{1, x, \ldots\}$ are linearly independent.
(c) $\{\sin kx, \cos kx \mid 1 \leq k \leq \ell\}$ are linearly independent $\{\sin kx, \cos kx \mid 1 \leq k\}$ are linearly independent. This is important for Fourier analysis.
(d) $\{1, c^x, c^{2x}, ...\}$ is linearly independent for $c > 1$.
(e) $\{1, c_0^x, c_1^x, ...\}$ is linearly independent where $c_0, c_1, ...$ are distinct constants not $0, 1$.

The above is also true when the functions have domain $\mathbb{N}$, i.e. whem considering the functions $\mathbb{N} \to \mathbb{R}$. Also, the above holds when the codomain is replaced by $\mathbb{C}$.

2.7 Basis and dimension

NOTE: If you like, you can restrict the following to finite bases. However there are infinite dimensional vector spaces. The simplest example is the set of power series (see previous section for other example).

A set B of vectors of V which are linearly independent and spans V is called a K–**basis** of V. All bases of V has the same size. Therefore we can define the K–**dimension** of V to be the size of any K–basis of V. This is denoted by basis dimension

$$\dim_K V$$

K–vector spaces are determined by their K–dimension, i.e., if V and V' are K–vector spaces, then

$$V \simeq V' \iff \dim_K V = \dim_K V'$$

where $\simeq$ denotes isomorphism of K–vector spaces. (This is the same even when the dimension is infinite.)

Note that since, when the field of scalars K is fixed, a K–vector space is determined by its dimension, if I say V is a K–vector space of dimension d, then V is structurally speaking the same as (i.e., is isomorphic to) the K–vector space K^d, the column (or row) vectors of size d. This gives the elements of V a familiar form for computations. For instance suppose V has dimension n and $B = \{e_0, ..., e_{n-1}\}$ is a fixed basis. Then each vector $v \in V$ can be written uniquely in the form

$$v = v_0 e_0 + \ldots + v_{n-1} e_{n-1}$$

where $v_i \in K$. This allows us to map V to K^n by

$$v_0 e_0 + \ldots + v_{n-1} e_{n-1} \mapsto \begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}$$

using column vector notation. (Or one can use row vector notation.) I will occasionally write

$$v_B = (v_0 e_0 + \ldots + v_{n-1} e_{n-1})_B = \begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}$$

if $v = v_0 e_0 + \ldots + v_{n-1} e_{n-1}$ where $B = \{e_0, ..., e_{n-1}\}$ This map $(\bullet)_B : V \to K^n$ is in fact an isomorphism of K–vector spaces. Note that $(\bullet)_B$ depends on B. To emphasize that the column vector is formed using B, one can write

$$\begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}_B$$

This subscript is only a reminder.

At this point, we have a way to represent vectors of V as column vectors:

$$(\bullet)_B : V \to K^n$$

depending on a choice of a basis B. What about linear maps?

If $f : V \to V'$ is a homomorphism of K–vector spaces where V has dimension c and V' has dimension r, both with elements viewed as column vectors, function f can be represented as a matrix multiplication. This of course assume the input $v \in V$ is written in column vector form for a choice of basis for V, say $B = \{e_0, ..., e_{c-1}\}$ and the output $f(v)$ is written as a column vector, also wrt to a chosen basis for V', say $B' = \{e'_0, ..., e'_{r-1}\}$. The size of this matrix is $r \times c$. This is possible because f is K–linear: Given a vector $v \in V$,

$$v = v_0 e_0 + \ldots + v_{c-1} e_{c-1}$$

where $e_0, ..., e_{c-1}$ is a basis of V, we have

$$f(v) = v_0 f(e_0) + \cdots + v_{c-1} f(e_{c-1})$$

Each $f(e_i)$ is a vector in V' and therefore

$$f(e_i) = f_{0,i} e'_0 + \cdots + f_{r-1,i} e'_{r-1}$$

where $e'_0, \ldots e'_{r-1}$ is a basis for V'. Therefore

$$\begin{aligned}
f(v) &= v_0 f(e_0) + \cdots + v_{c-1} f(e_{c-1}) \\
&= v_0(f_{0,0}e'_0 + \cdots + f_{r-1,0}e'_{r-1}) \\
&\quad + \ldots \\
&\quad + v_{c-1}(f_{0,c-1}e'_0 + \cdots + f_{r-1,c-1}e'_{r-1}) \\
&= (v_0 f_{0,0} + v_1 f_{0,1} + \cdots + v_{c-1} f_{0,c-1})e'_0 \\
&\quad + \ldots \\
&\quad + (v_0 f_{r-1,0} + v_1 f_{r-1,1} + \cdots + v_{c-1} f_{r-1,c-1})e'_0
\end{aligned}$$

Therefore the matrix is the computation

$$f(v_0 e_0 + ... + v_{c-1} e_{c-1})$$

where the input vector is expressed wrt B and the output vector is wrt B'

$$(f(v_0 e_0 + ... + v_{c-1} e_{c-1}))_{B'}$$

is the same as matrix multiplication

$$\begin{bmatrix} f_{0,0} & \cdots & f_{0,c-1} \\ \vdots & & \vdots \\ f_{r-1,0} & \cdots & f_{r-1,c-1} \end{bmatrix} \begin{bmatrix} v_0 \\ \vdots \\ v_{c-1} \end{bmatrix}$$

I will write

$$f_{B \to B'} = \begin{bmatrix} f_{0,0} & \cdots & f_{0,c-1} \\ \vdots & & \vdots \\ f_{r-1,0} & \cdots & f_{r-1,c-1} \end{bmatrix}$$

to indicate the matrix of f has a dependency on basis B (for input vectors) and B' (for output vectors). Therefore $(f)_{B \to B'}$ has three inputs:$f : V \to V'$, B, B'.

In the above our vector spaces V, V' each have their own fixed basis. There are times when you want to use different bases for the same vector space.

Suppose $B = \{e_i \mid 0 \leq i < n\}$ and $B' = \{e'_i \mid 0 \leq i < n\}$ are two bases of the same vector space V. Suppose I have a vector v that is expressed as column vector wrt basis B:

$$v_B = \begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}_B$$

and I want to re-expressed v using B':

$$v_{B'} = \begin{bmatrix} v'_0 \\ \vdots \\ v'_{n-1} \end{bmatrix}_{B'}$$

This can be achieved (just as before) using a matrix product.

$$v_{B'} = Av_B$$

where A is the matrix for **change of basis** from B to B'. AS a reminder, I might write $A_{B \to B'}$ for A. The above expression would then be (perhaps) easier to read:

change of basis

$$v_{B'} = A_{B \to B'} v_B$$

To get A, I will note write the linear transformation $f : V \to V$ and then harvest A as $(f)_{B \to B'}$. Now note that in this case f is the identity map $\mathrm{id} : V \to V$, $\mathrm{id}(v) = v$ since we are not changing the vector v – we are just re-expressing v using a different basis. Suppose

$$\mathrm{id}(e_i) = e_i = e_{i,0} e'_0 + \cdots + e_{i,n-1} e'_{n-1}$$

with $e_{i,j} \in K$. In other words I'm going to compute $(e_i)_{B'}$, the column vector of $e_i \in B$ wrt B'. Then $v = v_0 e_0 + \cdots + v_{n-1} e_{n-1}$ written in standard vector form using basis B can be re-expressed using basis B' as follows:

$$\begin{aligned}
\mathrm{id}(v) &= v \\
&= v_0 e_0 + \cdots + v_{n-1} e_{n-1} \\
&= v_0 \left(e_{0,0} e'_0 + \cdots + e_{n-1,0} e'_{n-1} \right) \\
&\quad + \cdots \\
&\quad + v_{n-1} \left(e_{0,n-1} e'_0 + \cdots + e_{n-1,n-1} e'_{n-1} \right) \\
&= \left(v_0 e_{0,0} + \cdots + v_{n-1} e_{n-1,0} \right) e'_0 \\
&\quad + \cdots \\
&\quad \left(v_0 e_{n-1,0} + \cdots + v_{n-1} e_{n-1,n-1} \right) e'_{n-1}
\end{aligned}$$

Expressing $\mathrm{id}(v)$ as a column vector wrt B':

$$(\mathrm{id}(v))_{B'} = \begin{bmatrix} e_{0,0} & \cdots & e_{n-1,0} \\ \vdots & & \vdots \\ e_{n-1,0} & \cdots & e_{n-1,n-1} \end{bmatrix} \begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}$$

with the understanding that the input v is written wrt B, i.e., $(v)_B = [v_0, ..., v_{n-1}]^t$

and the output vector is written wrt B'. Therefore the change of basis matrix from B to B' is

$$(\mathrm{id})_{B \to B'} = \begin{bmatrix} e_{0,0} & \cdots & e_{n-1,0} \\ \vdots & & \vdots \\ e_{n-1,0} & \cdots & e_{n-1,n-1} \end{bmatrix}$$

In terms of a matrix equation, if $(v)_B = [v_0, .., v_{n-1}]^t$ wrt B is converted to $(v')_B = [v'_0, .., v'_{n-1}]^t$ wrt B' the equation is

$$\begin{bmatrix} v'_0 \\ \vdots \\ v'_{n-1} \end{bmatrix}_{B'} = \begin{bmatrix} e_{0,0} & \cdots & e_{n-1,0} \\ \vdots & & \vdots \\ e_{n-1,0} & \cdots & e_{n-1,n-1} \end{bmatrix} \begin{bmatrix} v_0 \\ \vdots \\ v_{n-1} \end{bmatrix}_B$$

Again the subscripts are only reminders.

In linear algebra, it's common to use P and Q for change of basis matrices.

In the above, I change the column vector description of the same vector from one basis to another. Now consider the following.

Suppose I have a linear map $T : V \to V$. For computational purposes, I prefer to work with matrices. That means I need to fix a base B for V. Then the computation $T(v)$ can be done using

$$A(v)_B$$

where A is the matrix for T where v is now a column vector expressed using basis B, i.e,. $(v)_B$. But suppose I change my mind: I want to compute $T(v)$ but now I prefer to express the matrix of T assuming the input and output vectors are to be expressed using another basis B'. The question is this: how will A change?

The original computation was

$$A(v)_B$$

where A assumes that vectors $v \in V$ are expressed wrt B. Let $P = \mathrm{id}_{B \to B'}$ be the change of basis matrix from B to B'. The above is the same as

$$A(P^{-1}P)((v)_B) = (AP^{-1})(P(v)_B)$$

since $P^{-1}P = I$. The column vector $P(v)_B$ would then be v now expressed wrt B'. However the output is still written wrt B. To express the output using

B', you would do

$$P\left(AP^{-1}(P(v)_B)\right)$$

i.e.,

$$(PAP^{-1})(P(v)_B)$$

Of course since you now want to work exclusivly using B', your input would be wrt B' and not B. Therefore the change of basis on v expressed wrt B, $P(v)_B$, is not necessary. So we have

$$(PAP^{-1})(v)_{B'}$$

where the input $v_{B'}$ is vector v expressed as a column vector using B' (and not B). This has the same effect as T, but with the input expressed wrt B' and the output you get is also expressed wrt B' as well.

So the change of basis matrix from B to B' (the above P or the earlier notation $\mathrm{id}_{B\to B'}$) besides changing a vector wrt B to wrt B':

$$P(v)_B$$

can also change a matrix that works with basis B (for input and output) to another matrix of the same linear transformation that works with basis B' (for both input and output)

Note that the change of basis matrix $\mathrm{id}_{B\to B'}$ or P is invertible. (Because each e'_i is in the image of P, therefore $\dim imP \geq n$, which is in fact equality. Hence it is onto. By rank-nullity, the dimension of kernel is 0. Hence it is 1–1. All in all P is a bijection.)

The change of basis concept is very important:

Let A, A' be two (square) matrices of the same size. A and A' are **similar** matrices if there is an invertible matrix P such that $A' = P^{-1}AP$. If so, we write $A \sim A'$. Similarity is an equivalence relation: $A \sim A$, if $A \sim A'$, then $A' \sim A$, if $A \sim A'$ and $A' \sim A''$, then $A \sim A''$. similar

$A \sim A'$ means that the linear transformation that A, A' comes from are the same, other than the way vectors are expressed through a two different bases.

Many facts and information about a matrix A are preserved through similarity. For instance $\det(A) = \det(A')$ if A, A' are similar (see later). One would say that the determinant is invariant under similarity.

This appears in group theory in the following guise: In a group G, two elements

$x, y \in G$ are conjugates if there is some $g \in G$ such that $y = gxg^{-1}$. In fact there is no doubt that the conjugacy concept in group theory is borrowed from similarity in matrix theory.

2.8 Inner product

In $\mathbb{R}^n$, we can define dot product for $u = [u_0, ..., u_{n-1}]$ and $u = [v_0, ..., v_{n-1}]$,

$$u \cdot v = \sum_{i=0}^{n-1} u_i v_i$$

And when $n = 2$ or $n = 3$, the dot product is connected to the idea of angle θ between two vectors:

$$[u_0, ..., u_{n-1}] \cdot [v_0, ..., v_{n-1}] = |u||v| \cos\theta$$

where $|u| = \sqrt{\sum_{i=0}^{n-1} u_i^2}$. This angle allows us to define orthogonality: two vectors are **orthogonal** to each other if the angle between them is $\pi/2$. For general vector spaces, there's no natural way to define the angle between two vectors. However, we have the following fact that two vectors are orthogonal iff $u \cdot v = 0$. For a general vector space, the dot product is generalized to the concept of inner product:

orthogonal

Let V be a K–vector space where K is $\mathbb{R}$ or $\mathbb{C}$ (or more; see generalization below). Given $c \in K$, $\overline{c}$ is the complex conjugate of c. An **inner product** on V is a function

inner product

$$\langle \cdot, \cdot \rangle : V \times V \to K$$

such that

(a) Conjugate symmetry: $\overline{\langle u, v \rangle} = \langle v, u \rangle$
(b) Linearity in first component: $\langle cu + c'u', v \rangle = c\langle u, v \rangle + c'\langle u', v \rangle$
(c) Positive definiteness: If $v \neq 0$, then $\langle v, v \rangle > 0$.

An **inner product (vector) space** is a vector space with an inner product.

inner product (vector) space

Note that from conjugate symmetry, $\overline{\langle v, v \rangle} = \langle v, v \rangle$. Therefore $\langle v, v \rangle$ must be real. Also, $\langle 0, 0 \rangle = 0$ by linearity.

Also note that the linearity condition is only in the first component. This is sufficient since in most cases one would want to measure the contribution of a vector u along another vector v so that one would use $\langle u, v \rangle$. In particular if $e_0, ..., e_{n-1}$ is a basis for the vector space, one would compute $\langle u, e_i \rangle$ for $i = 0, ..., n-1$. So to understand say $u = c'u' + c''u''$, one might want to look at $\langle c'u' + c''u'', e_i \rangle$.

As for "linearity" in the second component:

$$\begin{aligned}
\langle u, cv + c'v' \rangle &= \overline{\langle cv + c'v', u \rangle} \\
&= \overline{c\langle v, u \rangle + c'\langle v', u \rangle} \\
&= \overline{c}\,\overline{\langle v, u \rangle} + \overline{c'}\,\overline{\langle v', u \rangle} \\
&= \overline{c}\langle u, v \rangle + \overline{c'}\langle u, v' \rangle
\end{aligned}$$

which is not quite linear. But if the field is $K = \mathbb{R}$, then $\overline{c} = c, \overline{c'} = c'$ and we do get linearity.

Note that, again relying on the dot product for $\mathbb{R}^2$ and $\mathbb{R}^3$, we have the concept of length:

$$|v| = \sqrt{v_0^2 + \ldots + v_{n-1}^2}$$

where $v = [v_0, \ldots, v_{n-1}]$. This is in fact

$$|v| = \sqrt{v \cdot v}$$

Therefore once a vector space has an inner product, we have an abstract concept of length.

Example 2.8.1. For $\mathbb{R}^n$, the dot product

$$u \cdot v = \sum_{i=0}^{n-1} u_i v_i$$

where $u = [u_0, \ldots, u_{n-1}]$ and $v = [v_0, \ldots, v_{n-1}]$ is an inner product. For $\mathbb{C}^n$,

$$u \cdot v = \sum_{i=0}^{n-1} u_i \overline{v_i}$$

is an inner product where $\overline{z}$ is the complex conjugate of z.

Example 2.8.2. Recall that $C^0([a, b])$ (for $a < b$) is the set of continuous functions $[a, b] \to \mathbb{R}$. Then

$$\langle f, g \rangle = \int_a^b f(x)g(x)\ dx$$

is an inner product for $C^0([a, b])$. If $C^0([a, b], \mathbb{C})$ is the set of continuous function $[a,b] \to \mathbb{C}$, then

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}\, dx$$

is an inner product for $C^0([a,b], \mathbb{C})$.

For a change of basis, suppose $B = \{e_0, ..., e_{n-1}\}$ and $B' = \{e'_0, ..., e'_{n-1}\}$ are two different bases for V, recall we have a change of basis matrix

$$\begin{bmatrix} e_{0,0} & \cdots & e_{n-1,0} \\ \vdots & & \vdots \\ e_{n-1,0} & \cdots & e_{n-1,n-1} \end{bmatrix}$$

where $e_{i,j}$ are defined by expressing every e_i as a linear combination of $e'_0, ..., e'_{n-1}$:

$$e_i = e_{i,0}e'_0 + \cdots + e_{i,n-1}e'_{n-1}$$

If V has an inner product $\langle \cdot, \cdot \rangle$ and B' is not just a basis but in fact is an orthogonal basis, i.e., e'_i, e'_j are orthogonal if $i \neq j$, then

$$\langle e_i, e'_j \rangle = \left\langle e_{i,0}e'_0 + \cdots + e_{i,n-1}e'_{n-1}, e'_j \right\rangle = e_{i,j}\langle e'_j, e'_j \rangle$$

i.e.

$$e_{i,j} = \frac{\langle e_i, e'_j \rangle}{\langle e'_j, e'_j \rangle}$$

Note that in the above computation, as I mentioned earlier, you only need linearity of an inner product for the first component. If in addition, if B' is a normal basis, i.e., $\langle e'_i, e'_i \rangle = 1$, then

$$e_{i,j} = \langle e_i, e'_j \rangle$$

Therefore an orthonormal basis B', a basis that is pairwise orthogonal and has element has length 1 where length is defined in terms of inner product

$$|v| = \sqrt{\langle v, v \rangle}$$

has a very clean change of basis matrix, i.e.,

$$\langle e'_i, e'_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Note that in the above, when I wrote

$$|v| = \sqrt{\langle v, v \rangle}$$

I'm using the fact that $\langle v, v\rangle \geq 0$.

Note that if I'm given a basis $B = \{e_0, e_1, e_2, ..., e_{n-1}\}$ I can adjust e_1 to make it orthogonal to e_0:

$$e_1' = e_1 - \frac{\langle e_1, e_0\rangle}{\langle e_0, e_0\rangle} e_0$$

The idea is that e_1 might be enroaching on e_0. Intuitively the term

$$\frac{\langle e_1, e_0\rangle}{\langle e_0, e_0\rangle} e_0$$

is the amount of e_1 in e_0.

Note that $e_0 \neq 0$ since it's a basis element. Therefore by positive definiteness, $\langle e_0, e_0\rangle \neq 0$. Now when I check if e_1' is orthogonal to e_0:

$$\begin{aligned}
\langle e_1', e_0\rangle &= \left\langle e_1 - \frac{\langle e_1, e_0\rangle}{\langle e_0, e_0\rangle} e_0, e_0 \right\rangle \\
&= \langle e_1, e_0\rangle - \frac{\langle e_1, e_0\rangle}{\langle e_0, e_0\rangle} \langle e_0, e_0\rangle \\
&= 0
\end{aligned}$$

The adjustment to e_2 is similar:

$$e_2' = e_2 - \frac{\langle e_2, e_0\rangle}{\langle e_0, e_0\rangle} e_0 - \frac{\langle e_2, e_1\rangle}{\langle e_1, e_1\rangle} e_1$$

etc. Therefore from any basis of an inner product space V, we can adjust the basis so that the new basis is an orthogonal basis. To make the basis orthonormal is easy: just normalize each element of the basis, i.e., divide it by it's length. (Remember a basis element cannot be 0 and therefore by positive definiteness, it's inner product with itself cannot be zero and hence the length is > 0.)

The above adjustment to a basis to make it orthogonal or orthonormal is call the Gram-Schmidt theorem:

Proposition 2.8.1. (Gram-Schmidt) *Let V be an inner product space.*

(a) *Let $\{e_0, e_1, ..., e_{k-1}\}$ is a set of linearly independent vectors. Define*

$$e_j' = e_j - \sum_{i<j} \frac{\langle e_j, e_i\rangle}{\langle e_i, e_i\rangle} e_i$$

for $j = 0, ..., k-1$*. Then* $\{e'_0, ..., e'_{k-1}\}$ *is a set of linearly independent and orthogonal vectors.*

(b) *Let* $\{e_0, e_1, ..., e_{n-1}\}$ *be a basis for* V*. Then* $\{e'_0, ..., e'_{n-1}\}$ *is an orthonormal basis for* V*.*

(c) *Let* $\{e_0, e_1, ..., e_{n-1}\}$ *be a basis for* V*. Then* $\{e'_0/|e'_0|, ..., e'_{n-1}/|e'_{n-1}|\}$ *is an orthonormal basis for* V*.*

2.9 Matrix

Recall from the previous section, if I let $(e_0, ..., e_{c-1})$ be a basis of V and $(e'_0, ..., e'_{r-1})$ be a basis of V'. If $f : V \to V'$ is a homomorphism of K–vectors spaces (i.e., a K–linear map), then

$$f(v_0 e_0 + \cdots v_{m-1} e_{c-1}) = v_0 f(e_0) + \cdots v_{m-1} f(e_{c-1})$$

Each

$$f(e_j)$$

is a vector in V'. Therefore

$$f(e_j) = f_{0,j} e'_0 + \cdots + f_{r-1,j} e'_{r-1}$$

Therefore f is fully determined by the $r \times c$ scalars

$$[f_{i,j}]_{0 \leq i < r,\ 0 \leq j < c}$$

This is the matrix of f wrt bases $(e_0, ..., e_{c-1})$ of V and $(e'_0, ..., e'_{r-1})$ of V'. The computation of the components of $f(v)$ can be computed using matrix product. If $v = v_0 e_0 + ... + v_{c-1} e_{c-1}$ then the values $v'_0, ..., v'_{r-1}$ where

$$f(v) = f(v_0 e_0 + ... + v_{c-1} e_{c-1}) = v'_0 e'_0 + \cdots + v'_{r-1} e'_{r-1}$$

is the matrix product

$$[f_{i,j}]_{0 \leq i < r,\ 0 \leq j < c} \cdot \begin{bmatrix} v_0 \\ \vdots \\ v_{c-1} \end{bmatrix}$$

So the function evaluation $f(v)$ is the same as matrix product computation where the vectors v and $f(v)$ are written as column vectors wrt fixed bases for V, V'.

If $f : V \to V'$ and $g : V' \to V''$ where V, V', V'' have dimensions p, q, r, then $g \circ f : V \to V''$ is also K–linear. If bases for V, V', V'' are fixed, then we have matrices $[f_{i,j}]_{0 \leq i < q,\ 0 \leq j < p}$ and $[g_{i,j}]_{0 \leq i < r,\ 0 \leq j < q}$ and the matrix of $g \circ f$ is the product of the matrices $[g_{i,j}]_{0 \leq i < r,\ 0 \leq j < q}$ and $[f_{i,j}]_{0 \leq i < q,\ 0 \leq j < p}$, i.e., the matrix of the composition $g \circ f$ is the product of the matrices of g and f.

Let $f : V \to V$ be a K–linear map. Then f is invertible (i.e., is an isomorphism) iff the matrix of f (when bases are fixed) $[f_{i,j}]_{0 \leq i < n,\ 0 \leq j < n}$ is an invertible matrix.

2.10 Elementary operations

One way to study a matrix A is to apply a sequence of simple matrices $E_0, ..., E_{k-1}$ to A so that you arrive at another simple matrix:

$$E_{k-1} \cdots E_0 A = B$$

This is usually done in stages: $E_0 A$, $E_1 E_0 A$, ... If E_i's are invertible, the above is essentially attempting to find a factorization of A:

$$A = (E_{k-1} \cdots E_0)^{-1} B = E_0^{-1} \cdots E_{k-1}^{-1} B$$

Even if the E_i's are not inverted, the above equation

$$E_{k-1} \cdots E_0 A = B$$

is still useful. This is because for some problems, one would like to associate to such a problem a matrix, say A. By choosing E_i carefully, the sequence $E_0 A$, $E_1 E_0 A$, ... are more or less the same as A wrt the problem, i.e., some invariant or some information is preserved.

The above is a very common and very important theme in the study of linear algebra and matrix theory.

It is because of the above, there are usually many "special" matrices. Examples are 0, I, diagonal matrices, elementary (row and column) matrices, (upper and lower) triangular matrices, orthogonal matrices, orthonormal matrices, etc.

The **elementary row operations** are operations performed on the rows of A: elementary row operations

(a) $R_i \leftrightarrow R_j$: Swap row i and row j.
(b) $R_i \to cR_i$: row i is replaced by a c multiple of row i (for $c \neq 0$).
(c) $R_i \to R_i + cR_j$: row i is replaced by the sum of row i and c multiple of row j (for $i \neq j$).

These row operations on a matrix can also be achieved through matrix multiplications.

I will use the following notation for the corresponding matrices of the above row operations: $R_{i,j}$ $R_i(c)$, $R_{i,j}(c)$. (This notation is mine and is not standard.) These matrices are obained by performing the operations on I_n where A is $n \times n$.

(a) $R_{i,j}$ is I_n with row i and row j swapped. Here's $R_{1,4}$ ($n = 7$); entries not shown are 0s:

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} & i & & & j & & \end{array} \\
\begin{array}{c} \\ i \\ \\ \\ j \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & \\
 & & & & 1 & & \\
 & & 1 & & & & \\
 & & & 1 & & & \\
 & 1 & & & & & \\
 & & & & & 1 & \\
 & & & & & & 1
\end{bmatrix}
\end{array}
$$

(b) $R_i(c)$ is I_n with row i replaced by c multiple of row i swapped. This is only used if $c \neq 0$. Here's $R_2(9)$ ($n = 7$):

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} & & i & & & & \end{array} \\
\begin{array}{c} \\ \\ i \\ \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & \\
 & 1 & & & & & \\
 & & 9 & & & & \\
 & & & 1 & & & \\
 & & & & 1 & & \\
 & & & & & 1 & \\
 & & & & & & 1
\end{bmatrix}
\end{array}
$$

(c) $R_{i,j}(c)$ is I_n with row i and replaced by the sum of row i and a c multiple of row j. This is only used when $c \neq 0$ and $i \neq j$. Here's $R_{2,5}(9)$ ($n = 7$):

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} & & i & & & j & \end{array} \\
\begin{array}{c} \\ \\ i \\ \\ \\ j \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & \\
 & 1 & & & & & \\
 & & 1 & & & 9 & \\
 & & & 1 & & & \\
 & & & & 1 & & \\
 & & & & & 1 & \\
 & & & & & & 1
\end{bmatrix}
\end{array}
$$

The following are obvious:

(a) $R_{i,j}^{-1} = R_{i,j}$
(b) $R_i(c)^{-1} = R_i(c^{-1})$
(c) $R_{i,j}(c)^{-1} = R_{i,j}(-c)$

There are corresponding elementary column operations.

For instance to compute the inverse of A

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

one can try to find $E_0, ..., E_{k-1}$ such that

$$E_{k-1}\cdots E_0 \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

For instance

$$\begin{aligned}
R_{1,0}(-3)\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} &= \begin{bmatrix}1 & 2\\ 0 & -2\end{bmatrix}\\
R_1(-1/2)\begin{bmatrix}1 & 2\\ 0 & -2\end{bmatrix} &= \begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}\\
R_{0,1}(-2)\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix} &= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}
\end{aligned}$$

i.e.,

$$(R_{0,1}(-2)R_1(-1/2)R_{1,0}(-3))\cdot\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

i.e.,

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}^{-1} = R_{0,1}(-2)R_1(-1/2)R_{1,0}(-3)$$

Instead of computing this product, equivalently, we can perform the corresponding row operations on I_2 since

$$R_{0,1}(-2)R_1(-1/2)R_{1,0}(-3) = R_{0,1}(-2)\left(R_1(-1/2)\left(R_{1,0}(-3)I_2\right)\right)$$

Therefore

$$\begin{aligned}
R_{0,1}(-2)R_1(-1/2)R_{1,0}(-3) &= R_{0,1}(-2)\left(R_1(-1/2)\left(R_{1,0}(-3)\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)\right)\\
&= R_{0,1}(-2)\left(R_1(-1/2)\begin{bmatrix}1 & 0\\ -3 & 1\end{bmatrix}\right)\\
&= R_{0,1}(-2)\left(\begin{bmatrix}1 & 0\\ 3/2 & -1/2\end{bmatrix}\right)\\
&= \begin{bmatrix}-2 & 1\\ 3/2 & -1/2\end{bmatrix}
\end{aligned}$$

Note that the row operations discovered to convert the given matrix A to I is applied in the same order to I in order to compute the inverse of the given matrix. Therefore, to cut down on writing, the usual paper computation is to apply the row operations to A and I at the same time:

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix} \xrightarrow{R_1 \to R_1 + (-3)\,R_0} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{bmatrix}$$
$$\xrightarrow{R_0 \to (-0.5)\,R_1} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1.5 & -0.5 \end{bmatrix}$$
$$\xrightarrow{R_0 \to R_0 + (-2)\,R_1} \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1.5 & -0.5 \end{bmatrix}$$

The pair of matrices

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

is called the **augmented matrix** of A. This is sometimes drawn with a vertical line in the middle to separate the two matrices:

augmented matrix

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right]$$

Hence A is invertible iff it is a product of elementary row operation matrices.

A matrix is in **row echelon form** up to row i if each row for row 0, ..., row i begins with a sequence of zeroes following possibly by nonzeroes (if any). Frequently this nonzero is 1, but it doesn't have to be. Any nonzero value would do. For such a row, the first nonzero value of 1 is called the **pivot** of that row. If $k < k'$ (where $0 \leq k < k' < i$) and there are pivots at row k and k', then the pivot at row k' is to the right of the pivot at row k. The following is in row echelon for row 0 and row 1:

row echelon form

pivot

$$\begin{bmatrix} 0 & \boxed{2} & ? & ? & ? & ? & ? \\ 0 & 0 & 0 & \boxed{\text{-3}} & ? & ? & ? \\ 0 & 0 & 0 & 0 & ? & ? & ? \\ 0 & 0 & 0 & 0 & ? & ? & ? \\ 0 & 0 & 0 & 0 & ? & ? & ? \end{bmatrix}$$

For the above matrix, to create (possibly) the third pivot, you look at the values to the immediate *right* and *below* the last pivot:

$$\begin{bmatrix} 0 & \boxed{2} & ? & ? & ? & ? & ? \\ 0 & 0 & 0 & \boxed{-3} & ? & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{?} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{?} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{?} & ? & ? \end{bmatrix}$$

If there's one:

$$\begin{bmatrix} 0 & \boxed{2} & ? & ? & ? & ? & ? \\ 0 & 0 & 0 & \boxed{-3} & ? & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{0} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{42} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{?} & ? & ? \end{bmatrix}$$

Perform a row swap so that row with the nonzero value is move to the top of these candidate rows:

$$\begin{bmatrix} 0 & \boxed{1} & ? & ? & ? & ? & ? \\ 0 & 0 & 0 & \boxed{1} & ? & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{42} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{0} & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{?} & ? & ? \end{bmatrix}$$

and you now have a new pivot value.

A matrix is in **row echolon form** if it is in row echolon for every row, such as this example:

row echolon form

$$\begin{bmatrix} 0 & \boxed{2} & ? & ? & ? & ? & ? \\ 0 & 0 & 0 & \boxed{-3} & ? & ? & ? \\ 0 & 0 & 0 & 0 & \boxed{42} & ? & ? \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Clearly using elementary row operations, one can transform a matrix to row echolon form. Assuming the matrix has size $m \times n$, the runtime is $O(mn \cdot \min(m, n))$.

A matrix is in **reduced row echelon form** up to row i if it is in row echelon form, each pivot is 1, and for each pivot up to row i, the values above the pivot

reduced row echelon form

are also zeroes. Here's an example:

$$\begin{bmatrix}
0 & \boxed{1} & ? & 0 & ? & ? & ? & 0 & ? & ? \\
0 & 0 & 0 & \boxed{1} & ? & ? & ? & 0 & ? & ? \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? & ? \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & ? & ? \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Another use of the elementary row operations is to simplify a system of linear equations.

2.11 Permutation matrices

Here's the matrix for $R_{1,3}$ for $n = 5$:

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(See previous section on elementary row operations.) If A is a $5 \times n$ matrix, then

$$EA$$

will swap row 1 and row 3 of B. If you have a collection of row swap matrices $E_0, ..., E_{k-1}$ of size n, then

$$E_{k-1} \cdots E_0$$

will be like I_n but with its rows permuted. Such a matrix is called a **permutation matrix**. Here's one:

permutation matrix

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

If you apply P to a matrix A (on the left), you'll get this:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} d & e & f \\ g & h & i \\ a & b & c \end{bmatrix}$$

The 1s in P are at column indices 1, 2, 0. Therefore PA is the matrix with rows row 1, row 2, row 0 of A. If you look at AP, you get

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} c & a & b \\ f & d & e \\ i & g & h \end{bmatrix}$$

Now P when applied to A (on the right), P permutes not the rows of A, but the columns of A. The 1s in the columns of P are at row indices 2, 0, 1. So AP is made up of column 2, column 0, column 1 of A.

There's also something very special about permutation matrices. If P is a permutation matrix, every row of P has exactly one 1 and every other value is 0 and every column of P has exactly one 1 and every other value is 0.

Furthermore, P^{-1} is really easy to compute:

$$P^{-1} = P^t$$

Permutation matrices are examples of **orthonormal matrices** in the sense that the rows of a permutation matrix form an orthonormal basis, i.e., it's a basis where every pair of vectors are orthogonal to each other ("ortho") and every vector has length 1 ("normal"). Furthemore the columns also form an orthonormal basis. If A is an orthogonal matrix, then orthonormal matrices

$$A^{-1} = A^t$$

(Proof: Easy.) An **orthogonal matrix** is one where the rows form an orthogonal basis (pairwise orthogonal) and might not have length 1. orthogonal matrix

Clearly the set of row permutation matrices forms a subgroup of $\mathrm{GL}_n(\mathbb{R})$ that is isomorphic (as groups) to the symmetric group on n symbols, S_n. Likewise for column permutation matrices.

Note that the set of row permutation matrices is the same as the set of column permutation matrices. The designation "row" and "column" technically speaking only applies when these matrices are applied to another matrix.

2.12 Other special matrices

A **diagonal matrix** is a square matrix where all off-diagonal entries are 0. Multiplying a diagonal matrix with another easy: the result is diagonal matrix where the (i, i)–entry is the product of the corresponding (i, i)–entries of the factors. Therefore powers are also easy to compute.

diagonal matrix

An **upper triangular matrix** is one where values strictly below the diagonal are 0. A **lower triangular matrix** is on where values strictly above the diagonal are 0. The product of two upper triangular matrices is upper triangular. The product of two lower triangular matrices is lower triangular. For both cases, the number of entries to compute is smaller.

upper triangular matrix
lower triangular matrix

An orthogonal matrix is a matrix where the columns (and rows) form an orthonormal basis. If A is an orthogonal matrix, then $A^tA = I = AA^t$. Hence $A^{-1} = A^t$. Therefore the inverse of A is easy to compute.

A **symmetric matrix** which is invariant (as a matrices) under transposition, i.e., A is symmetric is $A = A^t$.

symmetric matrix

Here are some examples of quadratic forms is

$$x^2 + xy + 2y^2$$

$$x^2 + y^2$$

$$x^2 + 7y^2 + z^2 + 3xz - 4xz$$

In the first two cases the quadratic forms involves x, y. The last involves x, y, z. Quadratic forms need not include all possible terms: in the third example above the xy term is missing. Let $p(x_0, ..., x_{n-1})$ be a quadratic form. p is **positive definite** if $p(x_0, ..., x_{n-1}) > 0$ for $(x_0, ..., x_{n-1})$ is a nonzero vector in $\mathbb{R}^n$.

positive definite

The study of quadratic forms appear in many areas of math, CS, physics, etc.

Note that

$$[x^2 + xy + 2y^2] = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Frequently in matrix theory, for a 1×1 matrix, by abuse, we consider that matrix as a scalar. (Think of [3.14159] as 3.14159.) So don't be surprise when you see

$$x^2 + xy + 2y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Note that A is symmetric.

Let A be an $n \times n$ symmetric matrix. If x is a column vector of n variables (say $x_0, ..., x_{n-1}$), then

$$x^t A x$$

is a quadratic form of n variables. The matrix A is said to **positive definite** if $x^t A x$ is a positive definite quadratic form. positive definite

2.13 Rowspace, columnspace, nullspace, and rank of a matrix

Let A be the matrix

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

The **column space** of A is the vector space spanned by the columns of A. This is denoted by $\text{colspace}(A)$. The **column rank** of A is the dimension of $\text{colspace}(A)$. This is denoted by $\text{colrank}(A)$.

column space

colspace(A) colrank

The **row space** of A is the vector space spanned by the rows of A. This is denoted by $\text{rowspace}(A)$. The **row rank** of A is the dimension of $\text{rowspace}(A)$. This is denoted by $\text{rowrank}(A)$.

row space

rowspace(A) rowrank

The above matrix A can be viewed as a function $\mathbb{R}^2 \to \mathbb{R}^2$ mapping vectors to vectors:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1x + 2y \\ 3x + 4y \end{bmatrix}$$

Likewise, if A is $r \times c$, then

$$A : \mathbb{R}^c \to \mathbb{R}^r$$

You can view the above matrix function this way

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1x + 2y \\ 3x + 4y \end{bmatrix} = \begin{bmatrix} 1x \\ 3x \end{bmatrix} + \begin{bmatrix} 2y \\ 4y \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} x + \begin{bmatrix} 2 \\ 4 \end{bmatrix} y$$

Therefore the image of A, $\operatorname{im} A$, is the linear combination of the columns of A, i.e., the column space of A, $\text{colspace}(A)$:

$$\operatorname{im}(A) = \text{colspace}(A)$$

and therefore

$$\dim \operatorname{im}(A) = \text{colrank}(A)$$

Note that in the above, *column* vectors are transformed by A through multiplying the vector on the *left* with A. One can also transform *row* vectors multiplying the vector on right by A. If I need to emphasize, if A is $r \times c$, I will write $A\bullet$ for the function that maps column vectors to column vectors:

$$A\bullet : \mathbb{R}^c \to \mathbb{R}^c$$

and

$$\bullet A : \mathbb{R}^r \to \mathbb{R}^c$$

for the function that maps row vectors to row vectors. Unless otherwise states, I will assume A is $A\bullet$.

Here's an example when A is not a square, say of size 2×3:

$$A : \mathbb{R}^3 \to \mathbb{R}^2$$

As before you can view the above matrix function this way:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1x + 2y + 3z \\ 2x + 4y + 6z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 4 \end{bmatrix} y + \begin{bmatrix} 3 \\ 6 \end{bmatrix} z$$

i.e., the vector input becomes coefficients for a linear combination of the columns of A, i.e., the image of A as a function is the span of the columns of A. Clearly for this example $\operatorname{im} A = \mathbb{R}^1$, i.e., $\dim \operatorname{im} A = 1$. Note that for the columns, we see the dependency of the second and third column on the first column as:

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} = c \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 6 \end{bmatrix} = d \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

where $c = 2, d = 3$. Putting this back into A, we see

$$\begin{bmatrix} 1 & c \cdot 1 & d \cdot 1 \\ 2 & c \cdot 2 & d \cdot 2 \end{bmatrix}$$

The rows of A are

$$\begin{bmatrix} 1 & c \cdot 1 & d \cdot 1 \end{bmatrix} \text{ and } \begin{bmatrix} 2 & c \cdot 2 & d \cdot 2 \end{bmatrix}$$

i.e., we now see the dependency of second row on the first row:

$$\begin{bmatrix} 1 & c \cdot 1 & d \cdot 1 \end{bmatrix} \text{ and } 2 \begin{bmatrix} 1 & c \cdot 1 & d \cdot 1 \end{bmatrix}$$

In other words $\dim \operatorname{rowspace}(A)$ is also 1. In general

$$\dim \operatorname{colspace}(A) = \dim \operatorname{rowspace}(A)$$

Theorem 2.13.1.

$$\operatorname{rowrank}(A) = \operatorname{colrank}(A) = \dim \operatorname{im}(A)$$

for any field.

(The row rank might not be the column rank over a division ring.) Therefore one can define the **rank** of a matrix to be either the row rank or the column rank of A. This is denoted by $\text{rank}(A)$. rank

Definition 2.13.1.

$$\text{rank}(A) = \text{rowrank}(A) = \text{colrank}(A) = \dim \text{im}(A)$$

The rank of A can be computed by performing row reduction to bring A to a row echelon form:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & 7 \\ 3 & 6 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

To find a basis for the column space, we bring it to the row reduced echelon form.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 3 & 6 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The following is a basis for the column space: Look for the 1s:

$$\begin{bmatrix} \boxed{1} & 2 & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \end{bmatrix}$$

A basis is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The row vectors

$$[1, 2, 0], \quad [0, 0, 1]$$

form a basis for the row space. There are two elements in both bases because the row rank and column rank has to be the same.

Let $A : \mathbb{R}^c \rightarrow \mathbb{R}^r$. The **kernel** of A is the set of vectors v such that kernel

$$Av = 0$$

This is denoted by $\ker A$. It's easy to show that $\ker A$ is a subspace of the domain vector space $\mathbb{R}^c$. This is also called the **null space** of A. This will also be denoted by $\text{nullspace}(A)$ The dimension of the null space of A is called the **nullity** of A. Here's an important fact: null space nullspace(A) nullity

Theorem 2.13.2. Rank-nullity theorem. *Let A be an $r \times c$ matrix, i.e., $A : \mathbb{R}^c \to \mathbb{R}^r$. Then* Rank-nullity theorem

$$\text{rank}(A) + \text{nullity}(A) = c = \dim(\text{dom}(A))$$

i.e.,

$$\dim(\text{im}(A)) + \dim\ker(A) = c = \dim(\text{dom}(A))$$

Proof.

$$\text{dom}(A)/\ker(A) \simeq \text{im}(A)$$

gives rise to the exact sequence

$$0 \to \ker(A) \to \text{dom}(A) \to \text{im}(A) \to 0$$

which splits for abelian categories so that

$$\text{dom}(A) \simeq \text{im}(A) \oplus \ker(A)$$

and hence

$$\dim\text{dom}(A) = \dim\text{im}(A) + \dim\ker(A)$$

□

For the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} : \mathbb{R}^3 \to \mathbb{R}^2$$

the function is

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1x + 2y + 3z \\ 2x + 4y + 6z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 4 \end{bmatrix} y + \begin{bmatrix} 3 \\ 6 \end{bmatrix} z$$

Clearly $\dim\text{colspace}(A) = 1$. Since $\dim\text{dom}(A) = 3$, by the rank-nullity theorem. $\dim\text{nullspace}(A) = 3 - 1 = 2$. Explicitly, $A \begin{bmatrix} x & y & z \end{bmatrix}^T = 0$ iff

$$\begin{bmatrix} 1x + 2y + 3z \\ 2x + 4y + 6z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e.,

$$1x + 2y + 3z = 0$$

i.e., the kernel of A is

$$\left\{ \begin{bmatrix} -2y - 3z & y & z \end{bmatrix}^T \mid y, z \in \mathbb{R} \right\}$$

which has dimension of 2.

2.14 Transpose

The transpose of a matrix is the matrix with its columns flipped so that column 0 becomes row 0, column 1 becomes row 1, etc. Formally if $A = [a_{i,j}]_{0 \leq i < m, 0 \leq j < n}$, the **transpose** of A is defined to be

transpose

$$A^t = [m^t_{i,j}]_{0 \leq i < n, 0 \leq j < m}$$

where

$$m^t_{i,j} = m_{j,i}$$

A matrix A is **symmetric** if $A^t = A$.

symmetric

Proposition 2.14.1.

(a) $I^t = I$
(b) $0^t = 0$
(c) $(AB)^t = B^t A^t$

(Some books write A^T or A^{T} instead of A^t.)

2.15 Inverse

If A is a square matrix of size $n \times n$, the **inverse** of A is matrix inverse

$$A^{-1}$$

such that

$$AA^{-1} = I_n = A^{-1}A$$

(To emphasize the defining property involves multiplication of matrices and not addition, one would say A^{-1} is the multiplicative inverse of A.)

Some matrices do not have inverses. A matrix is **invertible** if it has a (multiplicative) inverse. invertible

Multiplicative inverses are unique, i.e., if A has inverse A' and A'', i.e.,

$$AA' = I_n = A'A$$

and

$$AA'' = I_n = A''A$$

then $A' = A''$.

If A, B are invertible, then so is AB and

$$(AB)^{-1} = B^{-1}A^{-1}$$

2.16 Trace

The **trace** of a square matrix is the sum of diagonals of the matrix. Let $A = [a_{ij}]_{0 \leq i < n, 0 \leq j < n}$. Then trace

$$\operatorname{tr} A = \sum_{i=0}^{n-1} a_{ii}$$

Proposition 2.16.1. *Let A, B, C, P be square matrices of the same size. Furthermore assume P is invertible.*

(a) $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$
(b) $\operatorname{tr}(cA) = c \operatorname{tr}(A)$
(c) $\operatorname{tr} A = \operatorname{tr} A^t$
(d) $\operatorname{tr} A = \operatorname{tr} PAP^{-1}$
(e) $\operatorname{tr} AB = \operatorname{tr} BA$, $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$

2.17 Determinant

There are many equivalent definitions of the determinant of a square matrix. We will use this for our definition:

Definition 2.17.1. Let $A = [a_{ij}]_{0\leq i<n,\ 0\leq j<n}$ be a square matrix of size $n \times n$ with entries $a_{i,j}$ in field K. The **determinant** of A, denoted by determinant

$$\det(A)$$

is a scalar defined by the following. For convenience, if A is made up of columns vectors $v_0, ..., v_{n-1}$, we write $A = [v_0, v_1, ..., v_{n-1}]$. The determinant of A can be thought of as a function on the columns of A:

$$\det A = \det[v_0, v_1, ..., v_{n-1}] = \det(v_0, v_1, ..., v_{n-1})$$

(a) IDENTITY: If $A = I_n$, then $\det(A) = 1$,

(b) MULTILINEARITY: For any i such that $0 \leq i < n$, if $v_i = c'v'_i + c''v''_i$ where v', v'' are two column vectors and c', c'' are scalars, then

$$\det(v_0, ..., c'v'_i+c''v''_i, ...v_{n-1}) = c' \det(v_0, ..., v'_i, ...v_{n-1})+c'' \det(v_0, ..., v''_i, ...v_{n-1})$$

In other words $\det(A)$ is linear at every component when A is viewed as made up of column vectors:

$$\det(v_0, \ldots, v_{n-1}) = \det[v_0, ..., v_{n-1}]$$

We say det is a K–**multilinear** function in the sense that det is linear at every component $0 \leq i < n$. multilinear

(c) REPEAT COLUMNS: If A is written as a tuple of column vectors, if two columns are the same, then $\det(A) = 0$.

For instance

$$\begin{aligned}
&\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
&= 1 \det \begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix} + 3 \det \begin{bmatrix} 0 & 2 \\ 1 & 4 \end{bmatrix} \\
&= 1 \left(2 \det \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + 4 \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) + 3 \left(2 \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + 4 \det \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \right) \\
&= 1 \left(2 \cdot 0 + 4 \cdot 1\right) + 3 \left(2 \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + 4 \cdot 0 \right)
\end{aligned}$$

At this point, it's convenient to derive the following fact:

Proposition 2.17.1. SWAP COLUMNS:

$$\det[v_0, ..., v_i, ..., v_j, ..., v_{n-1}] = -\det[v_0, ..., v_j, ..., v_i, ..., v_{n-1}]$$

i.e., when two columns of a matrix are swapped the determinant changes sign.

(Proof: Exercise.)

Hence

$$\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = -1$$

Therefore, continuing the above calculation,

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = 1\,(2 \cdot 0 + 4 \cdot 1) + 3\,(2 \cdot (-1) + 4 \cdot 0) = 4 \cdot 1 + 3 \cdot 2 \cdot (-1) = -2$$

Proposition 2.17.2. *There is a unique function* $\mathrm{Mat}_{n,n}(K) \to K$ *that satisfies the above three properties.*

The above proposition is a useful tool for determinant computation. I'm going to list a few more:

Proposition 2.17.3. (a) SWAP COLUMNS:

$$\det[v_0, ..., v_i, ..., v_j, ..., v_{n-1}] = -\det[v_0, ..., v_j, ..., v_i, ..., v_{n-1}]$$

i.e., when two columns of a matrix are swapped the determinant changes sign.

(b) COLUMN MULTIPLICATION:

$$\det[v_0, ..., cv_i, ..., v_{n-1}] = c\det[v_0, ..., v_i, ..., v_{n-1}]$$

i.e., when a column of a matrix is multiplied by scalar c, the determinant is multiplied by c.

(c) ELEMENTARY COLUMN OPERATION:

$$\det[v_0, ..., v_i + cv_j, ..., v_{n+1}] = \det[v_0, ..., v_i, ..., v_{n+1}]$$

i.e., applying an elementary column operation $C_i \to C_i + cC_j$ to a matrix (where $i \neq j$) does not change its determinant.

Now it turns out the if $f : \mathrm{Mat}_{n,n}(K) \to K$ satisfies the conditions in the definition of det then f satisfies the properties in Proposition 2.17.3 (page 53). And if f satisfied the properties in Proposition 2.17.3 then f is the det function. Therefore we could have defined det using the properties in Proposition 2.17.3 and prove that such a det satisfies the properties in the definition.

In fact there are *many* equivalent definitions of the determinant function.

Here's the row version of the above proposition:

Proposition 2.17.4. *Let A be an $n \times n$ matrix. Viewing A as a tuple of n row vectors, $A = [v_0, ...v_{n-1}]$, we can think of $\det A$ as a function on the rows of A:*

$$\det A = \det(v_0, ...v_{n-1})$$

(a) Swap rows:

$$\det(v_0, ..., v_i, ..., v_j, ..., v_{n-1}) = -\det(v_0, ..., v_j, ..., v_i, ..., v_{n-1})$$

i.e., when two rows of a matrix are swapped the determinant changes sign.

(b) Row multiplication:

$$\det(v_0, ..., cv_i, ..., v_{n-1}) = c \det(v_0, ..., v_i, ..., v_{n-1})$$

i.e., when a row of a matrix is multiplied by scalar c, determinant is multiplied by c.

(c) Elementary row operation:

$$\det(v_0, ..., v_i + cv_j, ..., v_{n+1}) = \det(v_0, ..., v_i, ..., v_{n+1})$$

i.e., applying an elementary row operation $R_i \to R_i + cR_j$ to a matrix (where $i \neq j$) does not change its determinant.

It can be shown that f satisfies the conditions in the definition of det iff f satisfies the properties in Proposition 2.17.3 iff f satisfies the properties in Proposition 2.17.4.

It can be shown that the definition of det is equivalent to the following definition in the proposition below. We will need the following notation in the proposition: The (i, j)–**submatrix** of A, $\text{sub}_{i,j}(M)$, is the same as A but with row i and column j removed. If A is $n \times n$, then $\text{sub}_{i,j}(M)$ is $(n-1) \times (n-1)$. submatrix

(If A is a matrix, the (i, j)–submatrix is sometimes denoted by $A_{i,j}$. But I'm not using this notation because $A_{i,j}$ looks like an entry of matrix A.)

Proposition 2.17.5. Cofactor expansion by a row: *Define* $\det' : \text{Mat}_{n,n}(K) \to K$ *where*

$$\det'(A) = \sum_{j=0}^{n-1} (-1)^{i+j} a_{i,j} \det \text{sub}_{i,j}(A)$$

for any row i $(0 \leq i < n)$ *where* $a_{i,j}$ *is the* (i, j)*–entry of* A*. The above is recursive. The base case is when* A *is* 1×1*. In that case*

$$\det'[c] = c$$

where $[c]$ *is a* 1×1 *matrix with* $c \in K$*. Then* $\det A = \det' A$*.*

The expression

$$\det(A) = \sum_{j=0}^{n-1} (-1)^{i+j} m_{i,j} \det \text{sub}_{i,j}(A)$$

is call the **cofactor expansion** along row i. The cofactor expansion along any two rows will give the same value. The expression cofactor expansion

$$(-1)^{i+j} \det \text{sub}_{i,j}(A)$$

is called the (i, j)–**cofactor** of A while cofactor

$$\det \text{sub}_{i,j}(A)$$

is called the (i, j)–**minor** of A. minor

The definition of $\det, \det'$ is equivalent to the following:

Proposition 2.17.6. Cofactor expansion by a column: *Define* $\det''$: $\text{Mat}_{n,n}(K) \to K$ *where*

$$\det''(A) = \sum_{i=0}^{n-1} (-1)^{i+j} \det \text{sub}_{i,j}(A)$$

for any column j $(0 \leq j < n)$. *The above is recursive where* $n > 1$. *The base case is when* A *is* 1×1. *In that case*

$$\det''[c] = c$$

where $[c]$ *is a* 1×1 *matrix with* $c \in K$. *Then* $\det(A) = \det''(A)$.

In other words, $\det A = \det' A = \det'' A$.

There is yet another equivalent definition of the determinant using permutations:

Proposition 2.17.7. PERMUTATION: *Let* A *be* $n \times n$. *Define*

$$\det'''(A) = \sum_{\sigma \in S_n} \prod_{i=0}^{n-1} \operatorname{sgn}(\sigma) m_{i,\sigma(i)}$$

Then $\det'''(A) = \det(M)$.

The **signature** of a permutation, $\operatorname{sgn}(\sigma)$, is $(-1)^m$ where m is any number of transpositions when composed give us σ. signature

In the above definition of determinant using permutations, the product on the right hand side runs over row indices. You can also run over column indices and get the same value:

Proposition 2.17.8. *Let* A *be* $n \times n$. *Define*

$$\det''''(A) = \sum_{\sigma \in S_n} \prod_{j=0}^{n-1} \operatorname{sgn}(\sigma) m_{\sigma(j),j}$$

Then $\det''''(A) = \det(A)$.

The two computations of the determinant using permutations are the same because the signature of a permutation is the same as the signature of the inverse of the permutation and every element of S_n has a unique inverse.

Here is the computation of

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

using permutations. We have $S_2 = \{(0), (0,1)\}$ where $\operatorname{sgn}((0)) = 1$ and $\operatorname{sgn}((0,1)) = (-1)^1 = -1$. The determinant computations that runs over rows and columns are:

$$\begin{aligned}
\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} &= \sum_{\sigma \in \{(0),(0,1)\}} \prod_{i=0}^{n-1} \operatorname{sgn}(\sigma) a_{i,\sigma(i)} \\
&= \operatorname{sgn}((0)) a_{0,(0)(0)} a_{1,(0)(1)} + \operatorname{sgn}((0,1)) a_{0,(0,1)(0)} a_{1,(0,1)(1)} \\
&= 1 \cdot m_{0,0} m_{1,1} + (-1) \cdot m_{0,1} m_{1,0} \\
&= 1 \cdot (1)(4) + (-1) \cdot (2)(3) \\
&= -2 \\
\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} &= \sum_{\sigma \in \{(0),(0,1)\}} \prod_{j=0}^{n-1} \operatorname{sgn}(\sigma) a_{\sigma(j),j} \\
&= \operatorname{sgn}((0)) a_{(0)(0),0} a_{(0)(1),1} + \operatorname{sgn}((0,1)) a_{(0,1)(0),0} a_{(0,1)(1),1} \\
&= 1 \cdot a_{0,0} a_{1,1} + (-1) \cdot a_{1,0} a_{0,1} \\
&= 1 \cdot (1)(4) + (-1) \cdot (3)(2) \\
&= -2
\end{aligned}$$

So we have the following equivalent definitions

(a) Original definition where det is a function of column vectors
(b) Original definition where det is a function of row vectors
(c) Properties in Proposition 2.17.3 where input of det is column vectors
(d) Properties in Proposition 2.17.3 where input of det is row vectors
(e) Cofactor expansion by column
(f) Cofactor expansion by row
(g) Permutations of columns
(h) Permutations of rows

Here are some basic properties of the determinant function:

Proposition 2.17.9. *Let A, B, P be $n \times n$ matrices where P is invertible.*

(a) MULTIPLICATION: $\det(AB) = \det(A) \cdot \det(B)$
(b) TRANSPOSE: $\det(A^t) = \det(A)$
(c) CHANGE OF BASIS: $\det(PAP^{-1}) = \det(A)$
(d) INVERSE: *A is invertible iff $\det A \neq 0$. If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.*

2.18 System of linear equations

Because of the way matrix multiplication is defined, a system of linear equations

$$\begin{aligned} a_{0,0}x_0 + \cdots + a_{0,n-1}x_{n-1} &= b_0 \\ \vdots &= \vdots \\ a_{m-1,0}x_0 + \cdots + a_{m-1,n-1}x_{n-1} &= b_{m-1} \end{aligned}$$

corresponding to

$$Ax = b$$

where

$$A = \begin{bmatrix} a_{0,0} & \cdots & a_{0,n-1} \\ \vdots & & \vdots \\ a_{m-1,0} & \cdots & a_{m-1,n-1} \end{bmatrix}, \quad x = \begin{bmatrix} x_0 \\ \vdots \\ x_{m-1} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ \vdots \\ b_{m-1} \end{bmatrix}$$

We therefore have

$$Ax = b \implies PAx = Pb$$

where P is an $m \times m$ matrix. If P is invertible, we have

$$Ax = b \iff PAx = Pb$$

The goal is to find P such that $PAx = Pb$ is easier to work with than $Ax = b$.

In the case where A is a square, the perfect case is when A is invertible. In that case

$$Ax = b \iff x = A^{-1}b$$

$A^{-1}b$ can be computed through row reduction. Even if A is not a square, row reduction works. Therefore we will now continue with row reduction. If we arrive at this:

$$\left[\begin{array}{ccccccc|c} \boxed{1} & ? & 0 & ? & 0 & ? & ? & b'_0 \\ 0 & 0 & \boxed{1} & ? & 0 & ? & ? & b'_1 \\ 0 & 0 & 0 & 0 & \boxed{1} & ? & ? & b'_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b'_3 \end{array}\right]$$

putting in the x_i to make it easier to read:

$$\left[\begin{array}{ccccccc|c} \boxed{1}x_0 & ?x_1 & 0 & ?x_3 & 0 & ?x_5 & ?x_6 & b'_0 \\ 0 & 0 & \boxed{1}x_2 & ?x_3 & 0 & ?x_5 & ?x_6 & b'_1 \\ 0 & 0 & 0 & 0 & \boxed{1}x_4 & ?x_5 & ?x_6 & b'_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b'_3 \end{array}\right]$$

which means

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix} x_0 + \begin{bmatrix}0\\1\\0\\0\end{bmatrix} x_2 + \begin{bmatrix}0\\0\\1\\0\end{bmatrix} x_4 = \begin{bmatrix}-?\\0\\0\\0\end{bmatrix} x_1 + \begin{bmatrix}-?\\-?\\0\\0\end{bmatrix} x_3 + \begin{bmatrix}-?\\-?\\-?\\0\end{bmatrix} x_5 + \begin{bmatrix}-?\\-?\\-?\\0\end{bmatrix} x_6 + \begin{bmatrix}b'_0\\b'_1\\b'_2\\b'_3\end{bmatrix}$$

which means

$$\begin{aligned} x_0 &= -?x_1 + -?x_3 + -?x_5 + -?x_6 + b'_0 \\ x_2 &= -?x_3 + -?x_5 + -?x_6 + b'_1 \\ x_4 &= -?x_5 + -?x_6 + b'_2 \\ 0 &= b'_3 \end{aligned}$$

This means that if b'_3 is not zero, the system has no solutions. Equivalently, the original system has no solutions. The technical term is the last system and the first (since they are equivalent) are **inconsistent**. Otherwise, if $b'_3 = 0$, the solution set is 3-dimensional. Specifically if the reduce row echolon form (augmented) is

inconsistent

$$\left[\begin{array}{ccccccc|c} \boxed{1} & -3 & 0 & 2 & 0 & 7 & 1 & 3 \\ 0 & 0 & \boxed{1} & 0 & 0 & 3 & -1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 8 & -2 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Then

$$\begin{aligned} x_0 &= 3x_1 - 2x_3 - 7x_5 - x_6 + 3 \\ x_2 &= -5x_3 + 6x_5 + 1 \\ x_4 &= -8x_5 + 2x_6 + -7 \end{aligned}$$

where $x_1, x_3, x_5, x_6 \in \mathbb{R}$, or if you prefer to see all the x_i's on the left:

$$\begin{aligned}
x_0 &= 3r - 2s - 7t - u + 3 \\
x_1 &= r \\
x_2 &= -5s + 6t + 1 \\
x_3 &= s \\
x_4 &= -8t + 2u - 7 \\
x_5 &= t
\end{aligned}$$

where $r, s, t, u \in \mathbb{R}$. If you prefer a vector solution:

$$x = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} r + \begin{bmatrix} -2 \\ 0 \\ -5 \\ 1 \\ 0 \\ 0 \end{bmatrix} s + \begin{bmatrix} -7 \\ 0 \\ 6 \\ 0 \\ -8 \\ 1 \end{bmatrix} t + \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix} u + \begin{bmatrix} 3 \\ 0 \\ 1 \\ 0 \\ -7 \\ 0 \end{bmatrix}$$

where $r, s, t, u \in \mathbb{R}$ and $x = [x_0, x_1, ..., x_5]^t$.

Note that, ignoring b, let's look at the solution space $\ker A = \text{nullspace}(A)$ of x satisfying $Ax = 0$. From the reduced row echelon form, since there are 3 pivots,

$$\text{rank}(A) = 3$$

Also by rank-nullity theorem

$$\text{rank}(A) + \dim \text{nullspace} = \dim \text{dom}(A) = 7$$

Therefore

$$\dim \text{nullspace} = 4$$

And this is the case since the nullspace

The rank is due to the pivots:

$$\begin{array}{c}
\begin{array}{ccccccc} \downarrow & & \downarrow & & \downarrow & & \end{array} \\
\begin{bmatrix}
\boxed{1} & -3 & 0 & 2 & 0 & 7 & 1 \\
0 & 0 & \boxed{1} & 0 & 0 & 3 & -1 \\
0 & 0 & 0 & 0 & \boxed{1} & 8 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}$$

(or you can look at the pivots row-wise) and the nullity is due to the rest:

$$
\begin{array}{c}
\begin{array}{ccccccc}
\phantom{\boxed{1}} & \downarrow & & \downarrow & & \downarrow & \downarrow
\end{array} \\
\begin{bmatrix}
\boxed{1} & -3 & 0 & 2 & 0 & 7 & 1 \\
0 & 0 & \boxed{1} & 0 & 0 & 3 & -1 \\
0 & 0 & 0 & 0 & \boxed{1} & 8 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

From the computation of the pivots (and hence the rank), we already know, without computing the solutions explicitly, that the solutions will be parameterized by 4 variables, and hence the solutions of $Ax = 0$ when shifted by vector b will also involve 4 parameters.

Of course if the nullity is 0, the solution will be unique. If nullity is 0, then by rank-nullity theorem, the rank of A is the same as the dimension of the domain. An example would be

$$1x_0 = 42$$

i.e., $A = [1]$. But A need not be a square. For instance

$$
\begin{aligned}
1x_0 &= 42 \\
0x_1 &= 0
\end{aligned}
$$

i.e.,

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

also has rank 1.

2.19 LU and PLU decompositions

Recall that to solve this linear system

$$\begin{aligned} x + 2y &= 0 \\ 3x + 4y &= 1 \end{aligned}$$

we view this as a matrix equation

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

We then compute

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

for instance using elementary row operations.

We will now do the above in a different way. We will rewrite the matrix equation

$$Ax = c$$

into

$$LUx = c$$

i.e., we will rewrite matrix A into

$$LU = A$$

Here L is a **lower triangular matrix** and U is an **upper triangular matrix**. For $n = 2$, $LU = A$ looks like

lower triangular matrix
upper triangular matrix

$$\begin{bmatrix} 1 & \\ * & 1 \end{bmatrix} \begin{bmatrix} * & * \\ & * \end{bmatrix} = A$$

For $n = 3$, $LU = A$ looks like

$$\begin{bmatrix} 1 & & \\ * & 1 & \\ * & * & 1 \end{bmatrix} \begin{bmatrix} * & * & * \\ & * & * \\ & & * \end{bmatrix} = A$$

Entries not shown in the matrices are 0s. Note that for matrix L, the diagonal elements are 1s.

The factorization of $A = LU$ where L is lower triangular and U is upper triangular is called an **LU decomposition**.

LU decomposition

We can solve for L, U for the $n = 2$ case easily. From

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & \\ w & 1 \end{bmatrix} \begin{bmatrix} x & y \\ & z \end{bmatrix} = \begin{bmatrix} x & y \\ wx & wy + z \end{bmatrix}$$

we get

$$\begin{aligned} x &= a \\ y &= b \\ w &= c/a \\ z &= d - bc/a \end{aligned}$$

Clearly this works only if $a \neq 0$. For instance

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

Now let us solve a linear system of two equations in two unknowns using an LU decomposition. To solve

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

From the above, this is the same as

$$\begin{bmatrix} 1 & \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Now I do this: I split the above matrix equation into two equations. Let

$$\begin{bmatrix} 1 & \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 2 \\ & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix} \tag{2}$$

From (1) I get $x' = 0$ and $y' = -3x' + 1 = 1$. From (2) I get $y = y'/(-2) = -1/2$ and $x = -2y + x' = 1$.

The computation of L and U can also be done using elementary row operations, however the only elementary row operation you can use is $R_i \to R_i + cR_j$.

Here is an example for $n = 3$. Let me solve

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

but by first rewriting

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

as LU. We use the same idea as when I computed the inverse of a matrix. I use elementary row operations $E_0, ..., E_{k-1}$ only of the form $R_i \to R_i + cR_j$ $(i > j)$ such that

$$E_{k-1} \cdots E_0 \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} * & * & * \\ & * & * \\ & & * \end{bmatrix}$$

where the right hand side is an upper triangular matrix. Then we get

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = E_0^{-1} \cdots E_{k-1}^{-1} \begin{bmatrix} * & * & * \\ & * & * \\ & & * \end{bmatrix}$$

Note that if E_i is replacing row i with sum of row i and a c multiple of row j, $R_{j,i}(c)$ where $i < j$, then E_i is a lower rectangular matrix and so is its inverse.

$$\begin{aligned} R_{1,0}(-4) \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix} \\ R_{2,0}(-7) \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} \\ R_{2,1}(-2) \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Therefore

$$R_{2,1}(-2)R_{2,0}(-7)R_{1,0}(-4) \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

i.e.,

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} = R_{1,0}(-4)^{-1}R_{2,0}(-7)^{-1}R_{2,1}(-2)^{-1}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}$$

At this point we can multiply $R_{1,0}(-4)^{-1}R_{2,0}(-7)^{-1}R_{2,1}(-2)^{-1}$. However it's easier to perform row operations. Therefore we convert each of these terms to a row operation and then apply them to I_3:

$$\begin{aligned}
\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} &= R_{1,0}(-4)^{-1}R_{2,0}(-7)^{-1}R_{2,1}(-2)^{-1}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= R_{1,0}(4)R_{2,0}(7)R_{2,1}(2)\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= R_{1,0}(4)R_{2,0}(7)R_{2,1}(2)I_3\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= R_{1,0}(4)R_{2,0}(7)\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 2 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= R_{1,0}(4)\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\7 & 2 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= \begin{bmatrix}1 & 0 & 0\\4 & 1 & 0\\7 & 2 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}
\end{aligned}$$

(Aside: In the above computation of L, notice the following:

$$\begin{aligned}
\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} &= R_{1,0}(\boxed{\text{-4}})^{-1}R_{2,0}(\boxed{\text{-7}})^{-1}R_{2,1}(\boxed{\text{-2}})^{-1}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}\\
&= \ldots\\
&= \begin{bmatrix}1 & 0 & 0\\\boxed{4} & 1 & 0\\\boxed{7} & \boxed{2} & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\0 & -3 & -6\\0 & 0 & 0\end{bmatrix}
\end{aligned}$$

One can write less when computing L if one understands the above computation. Also, instead of collecting all the elementary row operations, reversing them, etc. one can start computing the L while going through the elementary operations. That's because the row elementary used commute with each other. To write less, one can do this:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 2 & 3 \\ 0 & 1 & 0 & 4 & 5 & 6 \\ 0 & 0 & 1 & 7 & 8 & 9 \end{array}\right] \xrightarrow{R_1 \to R_1 + (-4)\,R_0} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 2 & 3 \\ 4 & 1 & 0 & 0 & -3 & -3 \\ 0 & 0 & 1 & 7 & 8 & 9 \end{array}\right] \tag{1}$$

$$\cdots$$

$$\xrightarrow{R_2 \to R_2 + (-2)\,R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 2 & 3 \\ 4 & 1 & 0 & 0 & -3 & -3 \\ 7 & 2 & 1 & 0 & 0 & 0 \end{array}\right]$$

or simply omit the left submatrix and write down the L at the end of this sequence of row operations since each update to the left submatrix is independent of previous updates. Note that at (1), the left submatrix has a 4 and not a -4 because $R_1 \to R_1 + (\underline{-4})\,R_0$ is applied to the right submatrix but $R_1 \to R_1 + (\underline{4})\,R_0$ is applied to the left submatrix.)

Now we use the above LU decomposition to solve a system of linear equations. To solve

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

I use the LU decomposition just found and solve this:

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

I break this into two stage:

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{2}$$

From (1), we get $x' = 0$, $y' = -4x' + 1 = 1$, $z' = -7x' - 2y' + 2 = 0$. From (2), we get z has no constraint, $y = (6z + y')/(-3) = -2z - 1/3$, and $x = -2y - 3z + x' = -2(-2z - 1/3) - 3z = z + 2/3$. Let me write the above neatly: From (1) and (2):

$$
\begin{aligned}
x' &= 0 \\
y' &= (-4x' + 1) = 1 \\
z' &= -7x' - 2y' + 2 = 0 \\
z &= \text{any real number} \\
y &= (6z + y')/(-3) = -2z - 1/3 \\
x &= (-2y - 3z + x')/1 = z + 2/3 && (*)
\end{aligned}
$$

The solutions are

$$
\begin{aligned}
x &= z + 2/3 \\
y &= -2z - 1/3 \\
z &\in \mathbb{R}
\end{aligned}
$$

or in parametric vector form

$$
\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} t + \begin{bmatrix} 2/3 \\ -1/3 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}
$$

Note that the matrix L is not just a lower triangular matrix: it has 1s on the main diagonal. (That's because L is computed by applying $R_i \to R_i + cR_j$ $(i > j)$ to I.) To emphasize this, L is said to be a **lower unitriangular matrix**. Likewise if U has 1s on its main diagonal, U is said to be a **upper unitriangular matrix**.

lower unitriangular matrix
upper unitriangular matrix

Note that if A is either a lower or upper unitriangular matrix of size n, then $A - I_n$ is strictly lower or upper triangular, i.e., it's triangular and the main diagonal is made up of 0s. A strictly lower or upper triangular matrix A is **nilpotent**, i.e., there is some power $A^k = 0$. Clearly if that is the case, for $\ell \geq k$, $A^\ell = 0$. (More generally, in ring theory, an element r of a ring is nilpotent if there is some k such that $r^k = 0$. r is unipotent if $r - 1$ is nilpotent.) In particular if A is lower or upper unitriangular, $(A - I)^k = 0$ for $k \geq n$ where n is the size of A.

nilpotent

From the above LU decomposition

$$
\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 4 & 1 & \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ & -3 & -6 \\ & & 0 \end{bmatrix}
$$

I can make U an upper unitriangular matrix:

$$\begin{aligned}
\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} &= \begin{bmatrix} 1 & & \\ 4 & 1 & \\ 7 & 2 & 1 \end{bmatrix} R_1(-3)R_1(-1/3) \begin{bmatrix} 1 & 2 & 3 \\ & -3 & -6 \\ & & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 4 & 1 & \\ 7 & 2 & 1 \end{bmatrix} R_1(-3) \begin{bmatrix} 1 & 2 & 3 \\ & 1 & 2 \\ & & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 4 & 1 & \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -3 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ & 1 & 2 \\ & & 0 \end{bmatrix}
\end{aligned}$$

It's clear how to make an upper (or lower) triangular matrix into an upper (respectively lower) unitriangular matrix, except when a diagonal element is zero. The above decomposition of a square matrix A (if possible) is

$$A = LDU$$

where D is (clearly) a diagonal matrix.

Instead of solving a matrix equation $Ax = c$ using $LUx = c$ where $A = LU$ is the LU decomposition of A, we can also use the LDU decomposition of A.

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

we now break it down into three equations:

$$\begin{bmatrix} 1 & & \\ 4 & 1 & \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & & \\ & -3 & \\ & & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ & 1 & 2 \\ & & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{3}$$

From (1), we get $x'' = 0, y'' = -4x'' + 1 = 1, z'' = -7x'' - 2y'' + 2 = 0$. From (2), we get $x' = x'' = 0, y' = y''/(-3) = -1/3, z' = z'' = 0$. From (3), we get no constraint on z, $y = -2z + y' = -2z - 1/3$, $x = -2y - 3z + x' = -2(-2z - 1/3) - 3z + 0 = z + 2/3$, which is the same as the answer above.

If A has an LDU decomposition, then the LDU decomposition is unique. This is (obviously) not the case for LU decomposition.

The diagonal matrix D in an LDU decomposition factors out the scaling factors during Gaussian elimination process (or row reduction process). For special matrices, this is helpful in terms performance and storage. But for solving a general linear systems of equations using LDU decomposition is not necessarily better than using LU decomposition.

Whereas Gaussian elemination method and therefore more or less row elementary operations and row echolon form were more or less known for hundreds of years (although written using different notations), the idea of LU decomposition was discovered by astronomer Tadeusz Julian Banachiewicz only in 1938. Alan Turing formalized and further analyzed the LU decomposition in 1948 for the study of error propagation in numerical computations especially in matrix calculations. It was Turing who coined the term "LU decomposition" and used "L = lower" and "U = upper" notations for the two matrices in the decomposition.

Note that to solve a system of linear equations

$$Ax = b$$

you can use Gaussian elimination on the augmented matrix $[A|b]$. You can also compute the inverse of A^{-1} and the compute $x = A^{-1}b$. Or you can use the LU decomposition method. Gaussian elimination method is faster than computing the inverse matrix. There is one benefit to computing the inverse. Once the inverse A^{-1} is computed, you can store it and compute $A^{-1}b$ when necessary. This is the case when the b changes. The Gaussian elimination method is faster than the inverse matrix method but when b changes, you have to redo the Gaussian elimination calculation on $[A|b]$. However note that the inverse computation method does not work when the matrix is not invertible.

If you look at the earlier neatly written calculation (see $(*)$):

$$\begin{aligned}
x' &= 0 \\
y' &= (-4x' + 1) = 1 \\
z' &= -7x' - 2y' + 2 = 0 \\
z &= \text{any real number} \\
y &= (6z + y')/(-3) = -2z - 1/3 \\
x &= (-2y - 3z + x')/1 = z + 2/3
\end{aligned}$$

The general case is roughly this:

$$
\begin{aligned}
x' &= \boxed{?} \\
y' &= (?x' + \boxed{?}) \\
z' &= ?x' + ?y' + \boxed{?} \\
z &= z'/? \\
y &= (?z + y')/? \\
x &= (?y + ?z + x')/?
\end{aligned} \tag{1}
$$

where ? are from the L and U matrices and $\boxed{?}$ is from the right hand side column vector c. Note that for an equation such as (1), if the equation for that row in the matrix equation is $0z = ...$, there is no constraint on z. The values ? can be stored without recomputation when c is changed. The number of entries in L and U to stored (excluding the 0s and the 1s) is 8. In the case of the inverse matrix method, you have 9 values to store.

All in all, the LU method is much better, especially when there are many values of c to run over. This is in fact the reason Banachiewicz came up with the LU decomposition. He needed to solve a system of linear equations $Ax_0 = c_0$, $Ax_1 = c_1$, ..., $Ax_{n-1} = c_{n-1}$ in order to calculate astronomical and geodetic calculations of orbits where the right hand side column vector c_i are different but the left hand side A stays the same.

It is for the above reasons, most linear algebra computations in linear algebra libraries including

(a) solving a system of linear equations
(b) determinant computation (see below)
(c) inverse matrix computations (see below)

etc. are done using LU decomposition and its variations.

Now let us use the LU decomposition to compute determinants. Given

$$A = LU$$

we have

$$\det A = \det L \cdot \det U$$

This is then quickly computed the determinant of a triangular matrix (lower or upper) is the product of the elements in the diagonal. In fact since L is a unitriangular matrix, $\det L = 1$. Hence

$$\det A = \det U = \text{product of diagonal elements of } U$$

Next we will compute the inverse of a matrix using its LU decomposition. Assuming A is 3×3, we solve for $x_0 x_1, x_2$ are column vectors satisfying

$$Ax_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad Ax_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad Ax_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

which is the same as

$$A[x_0, x_1, x_2] = I_3$$

where $[x_0, x_1, x_2]$ is a matrix with columns x_0, x_1, x_2. Hence $[x_0, x_1, x_2]$ is the inverse of A.

Note that this is a case of solving several linear systems where the matrix A is the same but the right hand side column vectors are different.

Let us compute the inverse of

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

First I will compute the LU decomposition.

$$\begin{aligned}
R_{1,0}(-2) \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 3 & 2 & 1 \end{bmatrix} \\
R_{2,0}(-3) \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 3 & 2 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & -4 & -8 \end{bmatrix} \\
R_{2,1}(-4/3) \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & -4 & -8 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix}
\end{aligned}$$

Hence

$$R_{2,1}(-4/3) R_{2,0}(-3) R_{1,0}(-2) \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix}$$

Therefore

$$\begin{aligned}
\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} &= R_{1,0}(-2)^{-1} R_{2,0}(-3)^{-1} R_{2,1}(-4/3)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \\
&= R_{1,0}(2) R_{2,0}(3) R_{2,1}(4/3) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \\
&= R_{1,0}(2) R_{2,0}(3) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \\
&= R_{1,0}(2) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix}
\end{aligned}$$

To solve $Ax_0 = [1, 0, 0]^t$, i.e.,

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{02} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

we solve

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} x'_{00} \\ x'_{01} \\ x'_{02} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{02} \end{bmatrix} = \begin{bmatrix} x'_{00} \\ x'_{01} \\ x'_{02} \end{bmatrix} \tag{2}$$

From (1) and (2) we have

$$
\begin{aligned}
x'_{00} &= \boxed{1} \\
x'_{10} &= -2x'_{00} + \boxed{0} = 1 \\
x'_{20} &= -3x'_{00} - (4/3)x'_{10} + \boxed{0} = -1/3 \\
x_{20} &= x'_{20}/(-4) = \underline{1/12} \\
x_{10} &= (3x_{20} + x'_{10})/(-3) = \underline{7/12} \\
x_{00} &= -2x_{10} - 3x_{20} + x'_{00} = \underline{-5/12}
\end{aligned}
$$

The boxed values are the values of the right-hand side vector. Therefore

$$
x_0 = \begin{bmatrix} -5/12 \\ 7/12 \\ 1/12 \end{bmatrix}
$$

To solve $Ax_1 = [0,1,0]^t$, i.e.,

$$
\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix}
\begin{bmatrix} x_{01} \\ x_{11} \\ x_{21} \end{bmatrix}
=
\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}
$$

we solve

$$
\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix}
\begin{bmatrix} x'_{01} \\ x'_{11} \\ x'_{21} \end{bmatrix}
=
\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{1}
$$

$$
\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix}
\begin{bmatrix} x_{01} \\ x_{11} \\ x_{21} \end{bmatrix}
=
\begin{bmatrix} x'_{01} \\ x'_{11} \\ x'_{21} \end{bmatrix} \tag{2}
$$

From (1) and (2) we have

$$
\begin{aligned}
x'_{01} &= \boxed{0} \\
x'_{11} &= -2x'_{01} + \boxed{1} = 1 \\
x'_{21} &= -3x'_{01} - (4/3)x'_{11} + \boxed{0} = -4/3 \\
x_{21} &= x'_{21}/(-4) = \underline{1/3} \\
x_{11} &= (3x_{21} + x'_{11})/(-3) = \underline{-2/3} \\
x_{01} &= -2x_{11} - 3x_{21} + x'_{01} = \underline{1/3}
\end{aligned}
$$

Therefore

$$x_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

To solve $Ax_2 = [0,1,0]^t$, i.e.,

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_{02} \\ x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

we solve

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4/3 & 1 \end{bmatrix} \begin{bmatrix} x'_{01} \\ x'_{11} \\ x'_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_{02} \\ x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} x'_{02} \\ x'_{12} \\ x'_{22} \end{bmatrix} \tag{2}$$

From (1) and (2) we have

$$\begin{aligned}
x'_{02} &= \boxed{0} \\
x'_{12} &= -2x'_{02} + \boxed{0} = 0 \\
x'_{22} &= -3x'_{02} - (4/3)x'_{12} + \boxed{1} = 1 \\
x_{22} &= x'_{22}/(-4) = \underline{-1/4} \\
x_{12} &= (3x_{22} + x'_{12})/(-3) = \underline{1/4} \\
x_{02} &= -2x_{12} - 3x_{22} + x'_{02} = \underline{1/4}
\end{aligned}$$

Therefore

$$x_2 = \begin{bmatrix} 1/4 \\ 1/4 \\ -1/4 \end{bmatrix}$$

Putting

$$x_0 = \begin{bmatrix} -5/12 \\ 7/12 \\ 1/12 \end{bmatrix}, x_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}, x_2 = \begin{bmatrix} 1/4 \\ 1/4 \\ -1/4 \end{bmatrix}$$

together, we have

$$A^{-1} = \begin{bmatrix} -5/12 & 1/3 & 1/4 \\ 7/12 & -2/3 & 1/4 \\ 1/12 & 1/3 & -1/4 \end{bmatrix}$$

Note that removing the explanations, once L and U is computed and we have

$$\begin{aligned}
x'_{00} &= \boxed{1} \\
x'_{10} &= -2x'_{00} + \boxed{0} = 1 \\
x'_{20} &= -3x'_{00} - (4/3)x'_{10} + \boxed{0} = -1/3 \\
x_{20} &= x'_{20}/(-4) = \underline{1/12} \\
x_{10} &= (3x_{20} + x'_{10})/(-3) = \underline{7/12} \\
x_{00} &= -2x_{10} - 3x_{20} + x'_{00} = \underline{-5/12}
\end{aligned}$$

simplify change the values in the boxes with the values from x_1, x_2 and for each x_i, recompute the above equations and the last three underlined values is x_i. The equations for finding $x_i = [a, b, c]^t$ is then

$$\begin{aligned}
a' &= \boxed{1} \\
b' &= -a' + \boxed{0} \\
c' &= -3a' - (4/3)b' + \boxed{0} \\
c &= c'/(-4) \\
b &= (3c + b')/(-3) \\
a &= -2b - 3c + a'
\end{aligned}$$

where values in the boxes come from the column vector on the right.

In the above computation, at the earlier point where we have computed

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

we can compute the determinant:

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} = (1)(-3)(-4) = 12$$

In fact we only need U. Therefore we can already compute the determinant a little bit earlier without computing L: From

$$R_{2,1}(-4/3)R_{2,0}(-3)R_{1,0}(-2)\begin{bmatrix}1 & 2 & 3\\ 2 & 1 & 3\\ 3 & 2 & 1\end{bmatrix} = \begin{bmatrix}1 & 2 & 3\\ 0 & -3 & -3\\ 0 & 0 & -4\end{bmatrix}$$

we already have U (but not L), and therefore

$$\det\begin{bmatrix}1 & 2 & 3\\ 2 & 1 & 3\\ 3 & 2 & 1\end{bmatrix} = (1)(-3)(-4) = 12$$

We have three application of LU decomposition: solving equation (replacing the Gaussian elimination method), determinant computation, and inverse computation.

Note that while using elementary row operation to create an upper triangular matrix, you might come to a situation where the the next diagonal value is zero:

$$\begin{bmatrix}3 & * & * & * & * & *\\ & -2 & * & * & * & *\\ & & 4 & * & * & *\\ & & & \boxed{0} & * & *\\ & & & 9 & * & *\\ & & & 1 & * & *\end{bmatrix}$$

Remember that you cannot use the swap elementary row operation during LU decomposition. (If you do, the L will not be lower triangular.) This issue can be handled by performing elementary row operations of swapping rows on A before performing the LU decomposition. Letting P be the product of the matrices of these row swaps to ensure that the above situation(s) won't happen, we have PA and then we perform the LU decomposition:

$$PA = LU$$

Recall that such a matrix P is called a row permutation matrix. (See section on permutation matrices.) A column permutation matrix Q is defined in a similar way and permutes the columns of A by multiplication on the right side of A: AQ. Given matrix A, the expression

$$PA = LU$$

where P is a permutation matrix, L is lower triangular, and U is upper triangular is called a **PLU decomposition** of A. PLU decomposition

Here is an example. If we try to write

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

into LU, we immediately run into a problem. So we permute the row 0 and row 1 of A first:

$$\begin{aligned} A &= R_{0,1} R_{0,1} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 2 & 3 & 0 \end{bmatrix} \\ &= R_{0,1} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 3 & 0 \end{bmatrix} \\ &= R_{0,1} R_{2,0}(1) R_{2,0}(-1) \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 3 & 0 \end{bmatrix} \\ &= R_{0,1} R_{2,0}(1) \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \end{aligned}$$

Here, on the right, $R_{0,1}$ will contribute to P (on the left side of the equation), $R_{2,0}(1)$ will contribute to L, and remaining matrix remains to be decomposed.

Now when I pivot at row 1, I have the same problem. I need to swap row 1 and row 2. So I do this:

$$\begin{aligned} A &= \ldots \\ &= R_{0,1} R_{2,0}(1) \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \\ &= R_{0,1} R_{2,0}(1) R_{1,2} R_{1,2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \\ &= R_{0,1} R_{2,0}(1) R_{1,2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Now note that on the right, $R_{0,1}$ and $R_{1,2}$ are row swap operators. They form P a permutation matrix. The problem is $R_{2,1}(1)$, which contributes to L, is in the middle. We need to bring the row swap matrices together. I'm going to insert $R_{1,2}R_{1,2}$ like this:

$$
\begin{aligned}
A &= R_{0,1}R_{2,0}(1)R_{1,2}\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 0\end{bmatrix}\\
&= R_{0,1}\underline{R_{1,2}R_{1,2}}R_{2,0}(1)R_{1,2}\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 0\end{bmatrix}
\end{aligned}
$$

Then I get

$$
A = (R_{0,1}R_{1,2})\,(R_{1,2}R_{2,0}(1)R_{1,2})\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 0\end{bmatrix}
$$

Of course $R_{2,0}(1)$ is a lower triangular matrix:

$$
R_{2,0}(1) = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}
$$

But in fact $R_{1,2}R_{2,0}(1)R_{1,2}$ is also a lower triangular matrix:

$$
\begin{aligned}
R_{1,2}R_{2,0}(1)R_{1,2} &= R_{1,2}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}R_{1,2}\\
&= \begin{bmatrix}1 & 0 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}R_{1,2}\\
&= \begin{bmatrix}1 & 0 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}\\
&= \begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}
\end{aligned}
$$

Therefore

$$
A = (R_{0,1}R_{1,2})\begin{bmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 0\end{bmatrix}
$$

i.e.,

$$(R_{0,1}R_{1,2})^{-1}A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since

$$(R_{0,1}R_{1,2})^{-1} = R_{1,2}^{-1}R_{0,1}^{-1} = R_{1,2}R_{0,1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

we finally have

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

The above uses $R_{i,j}^2 = I$, $R_{i,j}(c)^{-1} = R_{i,j}(-c)$, and if L is lower triangle, then $R_{i,j}(c)LR_{i,j}(c)$ is also lower triangular.

Note that in solving $Ax = c$, the computation of P does not depend on c. Therefore for solving $Ax_i = c_i$ $(0 \leq i < n)$, the P is computed only once.

(a) For solving a system of linear equations, to solve $Ax = c$, solve $PAx = Pc$ by solving $LUx = Pc$. P is invertible (because P is a product of elementary row swapping matrices which are invertible). Therefore if $PAx = Pc$, we also have $Ax = c$.

(b) For determinant computation, recall that the $\det PA$ is the same as $\pm \det A$ where $\pm$ is determined by the parity number of row permutations. If there is an even number of row permutations, $\det PA = \det A$. If there is an odd number of row permutations, $\det PA = -\det A$. And of course $\det PA$ is computed using $\det U$.

(c) For inverse computations, once the inverse of PA is computed (using LU), from

$$(PA)^{-1} = A^{-1}P^{-1}$$

we have

$$A^{-1} = (PA)^{-1}P$$

where P should be viewed as a column permutation.

Of course LU and PLU decompositions are useful for other linear algebra computations (example: rank computation).

2.20 Eigenvalues and eigenvectors

Suppose you look at a linear transformation

$$T : V \to V'$$

say V, V' has dimensions n, m (resp). Fixing bases for V and V', we can study T as a matrix A:

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

where A is m-by-n. To make it simple, suppose $n = m$, i.e., A is square. It would be nice if T or A is "simple", i.e., easy to compute and easy to understand.

(In general A is determined by what it does to a basis. It would be really nice if A maps e_0 to a linear combination of e_0 instead of all the $e_0, ..., e_{n-1}$. Likewise for the other e_i. In that case A, say $Ae_i = \lambda_i e_i$, then A is just a diagonal matrix with values λ_0, λ_1,... λ_{n-1}. Of course you don't expect all linear maps to work like this.

In general if T is a linear map $V \to V'$, if there is a subspace $U \leq V$ then one can study $T|_U$. If $0 \to U \to V \to \text{im}(V) \to 0$ splits $V = U \oplus W$, which is the case for vector spaces (there are many proofs: vector spaces are modules over PID, category theory that vector spaces form an abelian category, etc.), then T as a linear transformation is $T = T|_U \oplus T|_W$, i.e., the matrix of T is made up of two diagonal blocks, i.e., there is a basis $e_0, ..., e_n$ that can be partitions into two disjoint subsets say $\{e_0, ..., e_{k-1}\}$, $\{e_k, ..., e_{n-1}\}$ such that T maps $\{e_0, ..., e_{k-1}\}$ to $\langle e_0, ..., e_{k-1}\rangle$ and $\{e_n, ..., e_{n-1}\}$ to $\langle e_k, ..., e_{n-1}\rangle$.)

This is not always possible. For instance look at this:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Then

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + y \\ y \end{bmatrix}$$

You can see information in the second dimension (the y) has "leaked" into the first.

Going back to the original idea: How would we find the U and V, equivalently find a basis for T to be made up of two diagonal blocks?

One way is to look at relations on T. Suppose

$$T^2 - 1 = 0$$

as a function where 1 means I_2 and 0 means the zero function or the zero matrix viewed as a function. Equivalently, with a choice of a basis, the matrix A satisfies

$$A^2 - I = 0$$

Back to $T^2 - 1 = 0$, We have

$$T^2 - 1 = (T - 1)(T + 1)$$

The question is then: If

$$(T - 1)(T + 1) = 0$$

then does this mean that $T - 1 = 0$ or $T + 1 = 0$. This is not the case since it's possible to find two matrices which are nonzero and yet their product is 0. But suppose we are in the situation where

$$(T - 1)(T + 1) = 0$$

then we are looking at the same situation in the study of polynomials. Given any polynomial $p(x)$ we are frequently interested in solving $p(x) = 0$ and we do that by factorizing $p(x)$ into simpler pieces:

$$p(x) = a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n)$$

and replace the original problem of $p(x) = 0$ by

$$a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) = 0$$

But let's take a leap of faith: Suppose it's possible to have $T - 1 = 0$. More generally suppose

$$(T - \lambda_1) = 0$$

Well we don't really expect that for the whole $\mathbb{R}^n$, i.e., we don't expect (or even most) T to behave like

$$T(v) = \lambda_1 v$$

for all v in the domain of T since this means that the matrix of T is a diagonal matrix with all diagonal elements λ_1. The polynomial factorization

$$a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) = 0$$

(if it exists) or in terms of T (or the matrix A of T),

$$a(T - \alpha_1)(T - \alpha_2) \cdots (T - \alpha_n) = 0$$

only means

$$a(T - \alpha_1)(T - \alpha_2) \cdots (T - \alpha_n)(v) = 0$$

i.e., there is some i such that So it's more reasonable to expect

$$a(T - \lambda_1)(v) = 0$$

i.e., on one dimension of $\mathbb{R}^n$, so let's say

$$(T - \lambda_1)(cv_1) = 0$$

for all $c \in \mathbb{R}$ and some fixed $v_1 \in \mathbb{R}^n$. To simplify things, we can just have

$$(T - \lambda_1)(v_1) = 0$$

(by linearity) for some nonzero v_1. So it seems that one thing we want to do is to find a vector v_1 and a number λ_1 that

$$Av_1 = \lambda v_1$$

If this is the case, then λ_1 is an eigenvalue of A and v_1 is the corresponding eigenvector.

Of course the question now is this: given a linear map how to we find a polynomial $p(x)$ such that

$$p(T) = 0$$

or if you prefer a matrix equation,

$$p(A) = 0$$

Look at

$$Av_0 = \lambda_0 v_0$$

from above, this is the same as

$$\lambda_0 v_0 - Av_0 = 0$$

which is the same as

$$(\lambda_0 I_n - A)v_0 = 0$$

Clearly we now need to look at

$$\det(\lambda I_n - A) = 0$$

which gives us a polynomial (in λ as a variable) of degree n. Once factorized, our λ_0 (and in fact the other λ's as well) will appear in the polynomial factorization.

That being said, remember that this involves factoring a polynomial of degree n, and possible over $\mathbb{R}$ (depending on the problem statement), you might get some quadratic irreducible factors. Of course over $\mathbb{C}$, we know from the fundamental theorem of algebra, then polynomial factors into linear terms. Don't forget that the irreducible factors (for $\mathbb{R}$ or for $\mathbb{C}$) might repeat.

[FROM ANOTHER SET OF NOTES] Let A be an $n \times n$ square matrix. If v is a nonzero vector and λ is a scalar such that

$$Av = \lambda v$$

we say that v is an **eigenvector** of λ. This means that we can calculate Av quickly. Furthermore, if v is an eigenvector for eigenvalue λ of A, eigenvector

$$A^n v = \lambda^n v$$

If v_i is an eigenvector for eigenvalue λ_i of A for $i = 0, ..., n-1$, then

$$A^n \sum_{i=0}^{n-1} c_i v_i = \sum_{i=0}^{n-1} c_i \lambda_i^n v_i$$

Hence if v is a vector in the domain of A and $\{v_i \mid 0 \leq i < n\}$ is a basis for the domain of A, by first expressing v as

$$v = \sum_{i=0}^{n-1} c_i v_i$$

one can compute

$$A^n v$$

quickly.

The **characteristic polynomial** of A is characteristic polynomial

$$p(x) = \det(xI_n - A)$$

$p(x)$ is monic with degree n. The roots of $p(x)$ are the **eigenvalues** of A. eigenvalues

An eigenvalue can be zero, but an eigenvector must be nonzero.

Let us try the above with

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

We have

$$0 = \det(\lambda I - A) = \det \begin{bmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 4 \end{bmatrix} = \lambda^2 - 5\lambda - 2$$

Therefore

$$\lambda = \frac{5 \pm \sqrt{33}}{2}$$

Let $\lambda_0 = (5+\sqrt{33})/2$ and $\lambda_1 = (5-\sqrt{33})/2$. (By the way, if the eigenvalues are real, by convention, it's common to list them in descending order.)

For λ_0, using $(\lambda_0 I - A)v_0 = 0$,

$$\begin{bmatrix} (5+\sqrt{33})/2 - 1 & -2 \\ -3 & (5+\sqrt{33})/2 - 4 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e.,

$$\begin{bmatrix} (3+\sqrt{33})/2 & -2 \\ -3 & (-3+\sqrt{33})/2 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives two relations. We just need one:

$$-3x_0 + \frac{-3+\sqrt{33}}{2} y_0 = 0$$

i.e.,

$$x_0 = \frac{-3+\sqrt{33}}{6} y_0$$

Hence

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \dfrac{-3+\sqrt{33}}{6} t_0 \\ t_0 \end{bmatrix}$$

where $t_0 \in \mathbb{R}$. Choosing $t_0 = 1$,

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \dfrac{-3+\sqrt{33}}{6} \\ 1 \end{bmatrix}$$

Let's test it:

$$Av_0 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} \dfrac{-3+\sqrt{33}}{6} \\ 1 \end{bmatrix} = \begin{bmatrix} 1\dfrac{-3+\sqrt{33}}{6} + 2 \\ 3\dfrac{-3+\sqrt{33}}{6} + 4 \end{bmatrix} = \begin{bmatrix} \dfrac{9+\sqrt{33}}{6} \\ \dfrac{5+\sqrt{33}}{2} \end{bmatrix}$$

and

$$\lambda_0 v_0 = \frac{5+\sqrt{33}}{2} \begin{bmatrix} \dfrac{-3+\sqrt{33}}{6} \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{5+\sqrt{33}}{2} \cdot \dfrac{-3+\sqrt{33}}{6} \\ \dfrac{5+\sqrt{33}}{2} \end{bmatrix} = \begin{bmatrix} \dfrac{9+\sqrt{33}}{2} \\ \dfrac{5+\sqrt{33}}{2} \end{bmatrix}$$

For λ_1, using $(\lambda_1 I - A)v_1 = 0$,

$$\begin{bmatrix} (5-\sqrt{33})/2 - 1 & -2 \\ -3 & (5-\sqrt{33})/2 - 4 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e.,

$$\begin{bmatrix} (3-\sqrt{33})/2 & -2 \\ -3 & (-3-\sqrt{33})/2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives two relations. We just need one:

$$-3x_1 + \frac{-3-\sqrt{33}}{2} y_1 = 0$$

i.e.,

$$x_1 = \frac{-3-\sqrt{33}}{6} y_1$$

Hence

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \dfrac{-3-\sqrt{33}}{6} t_1 \\ t_1 \end{bmatrix}$$

where $t_1 \in \mathbb{R}$. Choosing $t_1 = 1$:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \dfrac{-3-\sqrt{33}}{6} \\ 1 \end{bmatrix}$$

Let's test it:

$$Av_1 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} \dfrac{-3-\sqrt{33}}{6} \\ 1 \end{bmatrix} = \begin{bmatrix} 1\dfrac{-3-\sqrt{33}}{6} + 2 \\ 3\dfrac{-3-\sqrt{33}}{6} + 4 \end{bmatrix} = \begin{bmatrix} \dfrac{9-\sqrt{33}}{6} \\ \dfrac{5-\sqrt{33}}{2} \end{bmatrix}$$

and

$$\lambda_1 v_1 = \frac{5-\sqrt{33}}{2} \begin{bmatrix} \dfrac{-3-\sqrt{33}}{6} \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{5-\sqrt{33}}{2} \cdot \dfrac{-3-\sqrt{33}}{6} \\ \dfrac{5-\sqrt{33}}{2} \end{bmatrix} = \begin{bmatrix} \dfrac{9-\sqrt{33}}{2} \\ \dfrac{5-\sqrt{33}}{2} \end{bmatrix}$$

We have now found the eigenvectors of our λ_0, λ_1. What I'm going to do now is form a new matrix using the eigenvectors as column vectors:

$$P = \begin{bmatrix} v_0 & v_1 \end{bmatrix} = \begin{bmatrix} \dfrac{-3+\sqrt{33}}{6} & \dfrac{-3-\sqrt{33}}{6} \\ 1 & 1 \end{bmatrix}$$

The determinant is

$$\frac{-3+\sqrt{33}}{6} - \frac{-3-\sqrt{33}}{6} = \frac{\sqrt{33}}{3}$$

The inverse of P is

$$P^{-1} = \frac{3}{\sqrt{33}} \begin{bmatrix} 1 & -\dfrac{-3-\sqrt{33}}{6} \\ -1 & \dfrac{-3+\sqrt{33}}{6} \end{bmatrix} = \frac{\sqrt{33}}{11} \begin{bmatrix} 1 & \dfrac{3+\sqrt{33}}{6} \\ -1 & \dfrac{-3+\sqrt{33}}{6} \end{bmatrix}$$

Finally I'm going to compute $P^{-1}AP$:

$$
\begin{aligned}
P^{-1}AP &= P^{-1}\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}\begin{bmatrix}\dfrac{-3+\sqrt{33}}{6} & \dfrac{-3-\sqrt{33}}{6}\\ 1 & 1\end{bmatrix}\\
&= \frac{\sqrt{33}}{11}\begin{bmatrix}1 & \dfrac{3+\sqrt{33}}{6}\\ -1 & \dfrac{-3+\sqrt{33}}{6}\end{bmatrix}\begin{bmatrix}\dfrac{9+\sqrt{33}}{6} & \dfrac{9-\sqrt{33}}{6}\\ \dfrac{5+\sqrt{33}}{2} & \dfrac{5-\sqrt{33}}{2}\end{bmatrix}\\
&= \frac{\sqrt{33}}{11}\begin{bmatrix}\dfrac{33+5\sqrt{33}}{6} & 0\\ 0 & \dfrac{-33+5\sqrt{33}}{6}\end{bmatrix}\\
&= \begin{bmatrix}\dfrac{\sqrt{33}}{11}\cdot\dfrac{33+5\sqrt{33}}{6} & 0\\ 0 & \dfrac{\sqrt{33}}{11}\cdot\dfrac{-33+5\sqrt{33}}{6}\end{bmatrix}\\
&= \begin{bmatrix}\dfrac{33\cdot 5+33\sqrt{33}}{6\cdot 11} & 0\\ 0 & \dfrac{33\cdot 5-33\sqrt{33}}{6\cdot 11}\end{bmatrix}\\
&= \begin{bmatrix}\dfrac{5+\sqrt{33}}{2} & 0\\ 0 & \dfrac{5-\sqrt{33}}{2}\end{bmatrix}\\
&= \begin{bmatrix}\lambda_0 & 0\\ 0 & \lambda_1\end{bmatrix}
\end{aligned}
$$

We have just achieved the diagonalization of A, i.e., we found some matrix P such that $P^{-1}AP$ is diagonal. The fascinating thing is that the diagonal contains the eigenvalues and the change of basis matrix P is made up of eigenvectors (correponding to the ordering of the eigenvalues). For this to happen, there must be n linearly independent eigenvectors:

Let A be an $n \times n$ matrix. A is diagonalizable iff A has n linearly independent eigenvectors. And if A has n eigenvalues $\lambda_0, .., \lambda_{n-1}$ and n linearly independent eigenvectors $v_0, ..., v_{n-1}$, then $P^{-1}AP$ is diagonal with diagonal values $\lambda_0, .., \lambda_{n-1}$ and the columns of P are $v_0, ..., v_{n-1}$.

The proof is not difficult.

First assume A has n eigenvalues $\lambda_0, ..., \lambda_{n-1}$ and n corresponding eigenvectors $v_0, ..., v_{n-1}$. Let P the matrix with the n eigenvectors as columns, $P = [v_0, ..., v_{n-1}]$. From the n column vectors,

$$Av_0, Av_1, ..., Av_{n-1} = \lambda_0 v_0, \lambda_1 v_1, ..., \lambda_{n-1} v_{n-1}$$

we for an $n \times n$ matrix where the columns are the above column vectors:

$$[xAv_0, ..., Av_{n-1}] = [\lambda_0 v_0, ..., \lambda_{n-1} v_{n-1}]$$

Now note that the left hand side is AP and the right hand side is DP where D is the diagonal matrix with diagonal entries $\lambda_0, ..., \lambda_{n-1}$. We have

$$AP = DP$$

Hence

$$P^{-1}AP = D$$

Note that P^{-1} exists since $v_0, ..., v_{n-1}$ are linearly independent. (i.e., the linear span of $v_0, ..., v_{n-1}$ is the column space of A. Since they are linearly independent, $\mathrm{rank}(P) = \dim \mathrm{colspace}(P) = n$. Hence P is invertible.)

Conversely if A is diagonalizable, there is an invertible P such that

$$P^{-1}AP = D$$

where D is diagonal. Hence

$$AP = PD$$

As above, let the columns of P be $v_0, ..., v_{n-1}$: $P = [v_0, ..., v_{n-1}]$. Then AP has columns $Av_0, ..., Av_{n-1}$. On the right, PD has columns $[\lambda_0 v_0, ..., \lambda_{n-1} v_{n-1}]$. Since $AP = PD$, their columns must be the same, i.e.,

$$Av_i = \lambda_i v_i$$

Hence λ_i is an eigenvector of A and v_i is an eigenvector of λ_i. Note that v_i is nonzero otherwise P has a 0 column and is not invertible. Furthermore $v_0, ..., v_{n-1}$ are linearly independent since they are the columns of P and P is invertible.

(Historically, it's common to use Q and Λ instead of P and D so that the decomposition look like $A = Q\Lambda Q^{-1}$.)

If A is diagonalizable with eigenvalues $\lambda_0, ..., \lambda_{n-1}$ then

$$\det A = \prod_{i=0}^{n-1} \lambda_i$$

Note that A need not be diagonalizable. For instance if

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

then

$$0 = \det(\lambda I - A) = \begin{bmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{bmatrix} = (\lambda - 1)^2$$

Hence λ_0 occurs twice (i.e., multiplicity is 2). If we try to find the eigenvector of $\lambda_0 = 1$:

$$(1I - A)(v) = 0$$

i.e.,

$$\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e. $x_0 \in \mathbb{R}$ and $y_0 = 0$. We have at most one nonzero independent eigenvector. Therefore we cannot diagonalize A.

On the other hand,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

have eigenvalues $1, 1$ and $(1I - A)v = 0$ gives us

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

i.e., $x_0, y_0 \in \mathbb{R}$. We can choose $x_0 = 1, y_0 = 0$. For x_1, y_1, I can choose $x_1 = 0, y_1 = 1$. In this case, I have two linearly independent eigenvectors for eigenvalue 1 (of multiplicity 2). Of course in this case ($A = I_2$), I can easily see that $[1, 0]^t$ and $[0, 1]^t$ are eigenvectors of $\lambda_0 = 1$ without going through the above calculations.

In general the characteristic polynomial of A looks like

$$p_A(\lambda) = (\lambda - \lambda_0)^{n_0} \cdots (\lambda - \lambda_{k-1})^{n_{k-1}}$$

where

$$\sum_{i=0}^{k-1} n_i = n$$

where n_i is the multiplicity of λ_i. In general for each λ_i, it might be possible to find a certain number of linearly independent eigenvectors for λ_i. The number of such linearly independent eigenvectors is $\leq n_i$. This number, m_i, is frequently called the **geometric multiplicity** of λ_i. To differentiate from n_i, the number n_i is frequently called the **algebraic multiplicity** of λ_i.

geometric multiplicity

algebraic multiplicity

example eigenvalue = 0

example geom mult = 1 < alg mult = 2

example geom mult = 2 = alg mult

2.21 Jordan canonical form

2.22 QR decomposition

2.23 Cholesky decomposition

2.24 Singular value decomposition

2.25 Principal component analysis

Index