Linear Algebra Lecture Notes
Math 2418 – Fall 2011

Dr. Richard Ketchersid

1 Introduction

We will study how to solve four central problems in Linear Algebra and the associated matrix decompositions that arise from the solutions

Linear Systems	$A\boldsymbol{x} = \boldsymbol{b}$	$n \times n$ matrix A	$A = LU$
Least Squares	$A\boldsymbol{x} = \boldsymbol{b}$	$m \times n$ matrix A	$A = QR$
Eigenvalues	$A\boldsymbol{x} = \lambda \boldsymbol{x}$	$n \times n$ matrix A	$A = PDP^{-1}$
Singular values	$A\boldsymbol{v} = \sigma \boldsymbol{u}$	$m \times n$ matrix A	$A = U\Sigma V^T$

1.1 Scalar Fields

We will primarily consider *real vector spaces*, but we will have occasion to consider *complex vector spaces*. The type of vector space is determined by the *field of scalars*, this will be either $\mathbb{R}$ or $\mathbb{C}$.

1.1.1 Review of complex numbers

Recall that the *complex numbers*, $\mathbb{C}$, is the set of numbers of the form $z = a + ib \in \mathbb{C}$ with $a, b \in \mathbb{R}$, where $i^2 = -1$, or if you like, $i = \sqrt{-1}$.

The *magnitude of a complex number*, $z = a + ib$, is defined to be

$$|z| \stackrel{\text{df}}{=} \sqrt{a^2 + b^2}$$

and the *conjugate* of $z = a + ib$ is

$$z^* \stackrel{\text{df}}{=} a - ib.$$

Notice

$$\boxed{z \text{ is real } \Leftrightarrow z^* = z},$$

also

$$\boxed{z^* z = |z|^2}$$

Since $z\left(\dfrac{z^*}{|z|^2}\right) = \dfrac{z^* z}{|z|^2} = \dfrac{|z|^2}{|z|^2} = 1$,

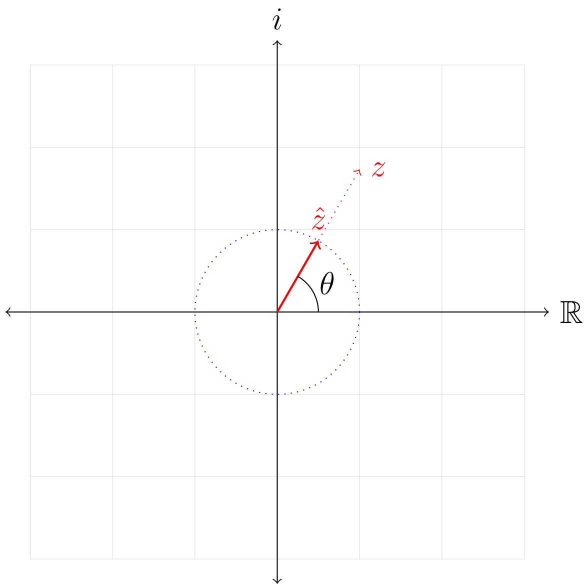
$$\boxed{\frac{1}{z} = z^{-1} = \frac{z^*}{|z|^2}}.$$

Problem 1 Write each of the following in the form $a + ib$:

(a) $\dfrac{1}{2 - 3i}$.

(b) $\dfrac{2+5i}{3+2i}$

For a complex number z let $\hat{z} = \frac{z}{|z|}$ so $|\hat{z}| = 1$, $\hat{z} = \cos(\theta) + i\sin(\theta)$ and $z = |z|\hat{z} = |z|(\cos(\theta) + i\sin(\theta))$.



Using the fact[1] that $\cos(\theta) + i\sin(\theta) = e^{i\theta}$ it is convenient to express a complex number as $z = re^{i\theta}$, where r is the magnitude of

[1] This is Euler's formula and can be derived from the infinite series representation of e^{it} as follows using the fact that $i^2 = -1$:

$$
\begin{aligned}
e^{it} &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} = 1 + it + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \frac{(it)^4}{4!} + \frac{(it)^5}{5!} + \cdots \\
&= \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots\right) + \left(it - i\frac{t^3}{3!} + \frac{t^5}{5!} + \cdots\right) \\
&= \sum_{n=0}^{\infty}(-1)^n \frac{t^{2n}}{(2n)!} + i\sum_{n=0}^{\infty}(-1)^n \frac{t^{2n+1}}{(2n+1)!} = \cot(t) + i\sin(t)
\end{aligned}
$$

z and $e^{i\theta}$ is the point on the unit circle indicating the direction.

Problem 2 Show that $z^* = re^{-i\theta}$ when $z = re^{i\theta}$, so that, $zz^* = (re^{i\theta})(re^{-i\theta}) = r^2 e^{i\theta - i\theta} = r^2 e^0 = r^2$.

Problem 3 Write z as $r(\cos(\theta) + i\sin(\theta))$ and as $re^{i\theta}$ for each of the following z:

(a) $z = \frac{3}{\sqrt{2}} - i\frac{3}{\sqrt{2}}$.

(b) $z = -1 + i\sqrt{3}$.

Write the following as $a + ib$:

(c) $3e^{i\frac{5\pi}{6}}$.

(d) $4(\cos\left(\frac{3\pi}{4}\right) + i\sin\left(3\pi\right)4)$.

Problem 4 Geometrically what does multiplication by $e^{i\beta}$ do? What does multiplication by a real α do?

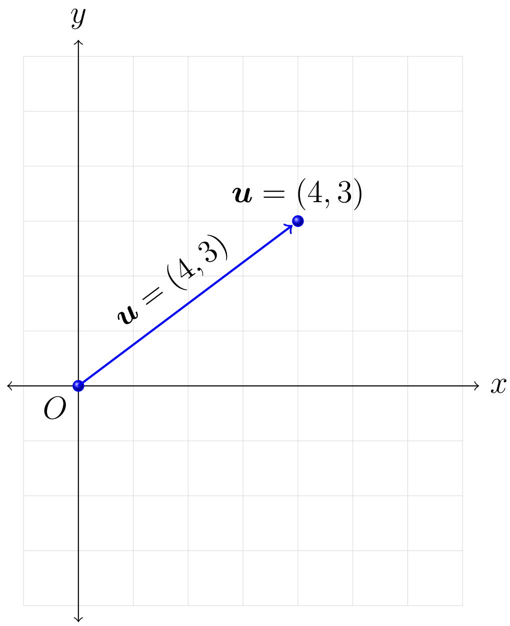
Problem 5 Using the representation $z = re^{i\theta}$ geometrically and algebraically describe the n, n^{th} roots of z. Which one should be called the *principle root*? What are the 3 cubed roots of -1? What is the principle cubed root of -1?

These are the basic facts to remember about complex numbers.

1.2 (Real) Euclidean vector spaces $\mathbb{R}^n$

Euclidean n-space, $\mathbb{R}^n$, is the set of all n-tuples of reals. An element, $\boldsymbol{x} = (x_1, \ldots, x_n)$, of $\mathbb{R}^n$ will be considered both as a *point* and as a *vector*, which can be viewed as a directed line segment

(arrow) from the origin to $\boldsymbol{x}$. Mathematically, we do not distinguish between points and vectors in $\mathbb{R}^n$.



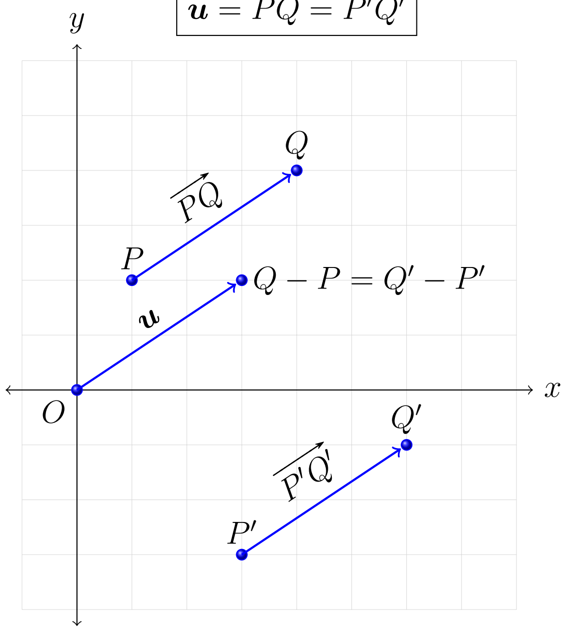
Notation We will denote elements of $\mathbb{R}^n$ as tuples and also as *column vectors*

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n) = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$$

For example $(1, 3, 2)$ in $\mathbb{R}^3$ can also be written as $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$. ♢

Points $P = (x_1, \ldots, x_n)$ and $Q = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ determine the vector $\overrightarrow{PQ}$ which could be viewed as having its tail at P and head at Q. We do not distinguish between vectors of the same *length* and *direction* and thus identify $\overrightarrow{PQ}$ with the corresponding vector $\boldsymbol{u} = Q - P = (y_1 - x_1, y_2 - x_2, \ldots, y_n - x_n)$ with tail at the origin.

$$\boxed{\boldsymbol{u} = \overrightarrow{PQ} = \overrightarrow{P'Q'}}$$



Clearly

$$\boxed{\overrightarrow{PQ} = \overrightarrow{P'Q'} \Leftrightarrow Q - P = Q' - P'}$$

Problem 6 Given that $P = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\boldsymbol{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, for what point Q would $\overrightarrow{PQ} = \boldsymbol{u}$? Draw the corresponding vectors and points.

1.2.1 Linear combinations

There is a natural algebraic structure on vectors. Given
$\boldsymbol{u} = (u_1, \ldots, u_n)$ and $\boldsymbol{v} = (v_1 \ldots, v_n)$ in $\mathbb{R}^n$ and scalar $\alpha \in \mathbb{R}$:

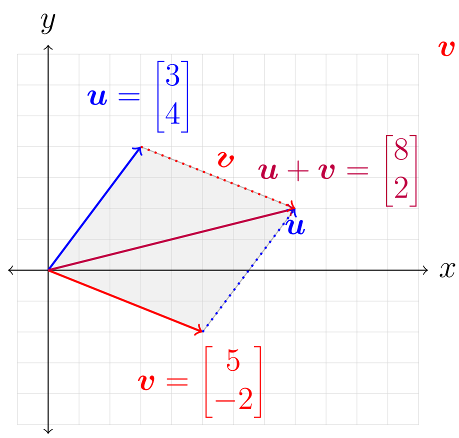
- (vector addition/component-wise addition)

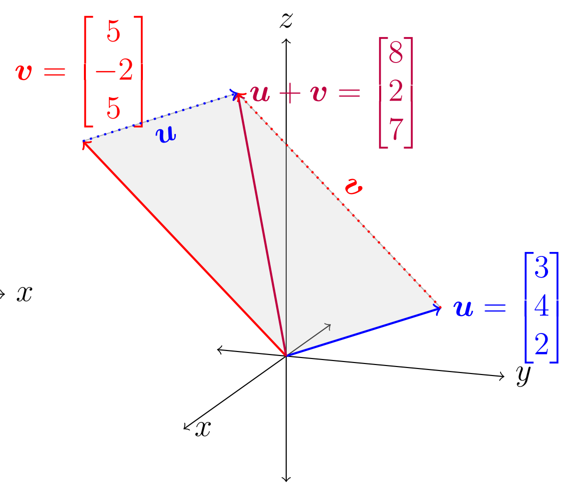
$$\boldsymbol{u}+\boldsymbol{v} = (u_1,\dots,u_n)+(v_1,\dots,v_n) = (u_1+v_1,\dots,u_n+v_n)$$
$$\text{or} \quad \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1+v_1 \\ \vdots \\ u_n+v_n \end{bmatrix}$$

- (scalar multiplication/component-wise multiplication)

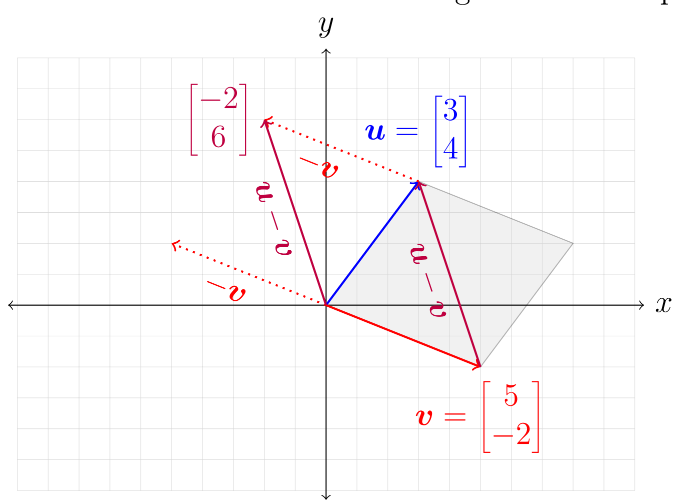
$$\alpha\boldsymbol{v} = \alpha(v_1,\dots,v_n) = (\alpha v_1,\dots,\alpha v_n)$$
$$\text{or } \alpha \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \alpha v_1 \\ \vdots \\ \alpha v_n \end{bmatrix}$$

The vector addition, $\boldsymbol{u}+\boldsymbol{v}$, can be viewed, geometrically, as finding the resultant vector when attaching $\boldsymbol{v}$ to the head of $\boldsymbol{u}$, or vice versa



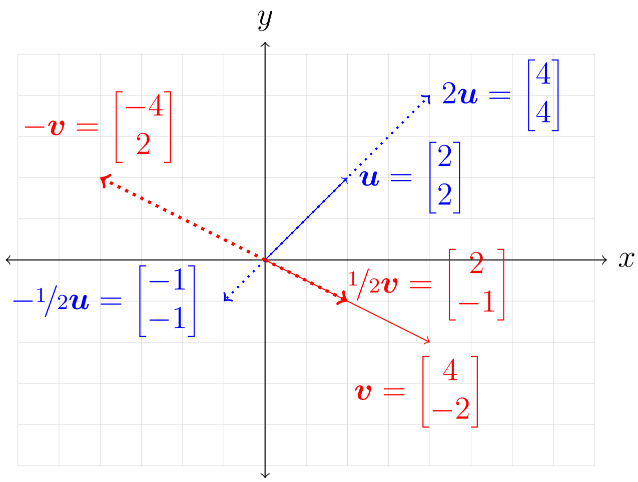


Vector subtraction has a similar geometric interpretation.



Notice that this agrees with our earlier use of $\overrightarrow{\boldsymbol{vu}} = \boldsymbol{u}-\boldsymbol{v}$.

Scalar multiplication can be viewed geometrically as scaling, stretching/dilating, and perhaps inverting direction, reflecting through the origin.



Important! A *linear combination of the vectors* $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ in $\mathbb{R}^n$ is

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots \alpha_k \boldsymbol{v}_k$$

where the α_i are scalars.

Problem 7 Write down 3 vectors in each of $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$ and find 2 different linear combinations of those vectors. Draw pictures for your your linear combinations in $\mathbb{R}^2$.

✐

Problem 8 Let $\boldsymbol{u} = (1, 2, 3)$, $\boldsymbol{v} = (-3, -2, -1)$, and $\boldsymbol{w} = (1, -1, 0)$.

(a) Find $2\boldsymbol{u} - 3\boldsymbol{w}$.

(b) Find scalars α, β, γ so that $\alpha\boldsymbol{u} + \beta\boldsymbol{v} + \gamma\boldsymbol{w} = (0, -2, 2)$.

(c) **Important Concept!** Describe algebraically and geometrically the following subsets of $\mathbb{R}^3$:

(i) All linear combinations of $\boldsymbol{u}$?

(ii) All linear combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$?

(iii) All linear combinations of $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$?

Problem 9 For $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^2$ which are not co-linear, describe geometrically the set of all linear combinations of $\boldsymbol{u}$ and the set of all linear combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$. What if $\boldsymbol{u}$ and $\boldsymbol{v}$ had come from $\mathbb{R}^3$?

Problem 10 For $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^3$ which are not coplanar, describe geometrically the set of all linear combinations $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$.

Problem 11 For what scalars c, d do the linear combinations $c\begin{bmatrix}1\\2\end{bmatrix} + d\begin{bmatrix}-1\\1\end{bmatrix}$ fill the triangle with vertices $(0,0)$, $(1,2)$, and $(-1,1)$.

1.2.2 Standard Basis

The i^{th} standard basis element for $\mathbb{R}^n$ is denoted $\boldsymbol{e}_i^n$ and is the element of $\mathbb{R}^n$ with a 1 in the i^{th} position and a 0 everywhere else. In $\mathbb{R}^2$ and $\mathbb{R}^3$ there is a tradition of using $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$.

In $\mathbb{R}^2$:

$$\boldsymbol{e}_1^2 = \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \boldsymbol{e}_2^2 = \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

In $\mathbb{R}^3$:

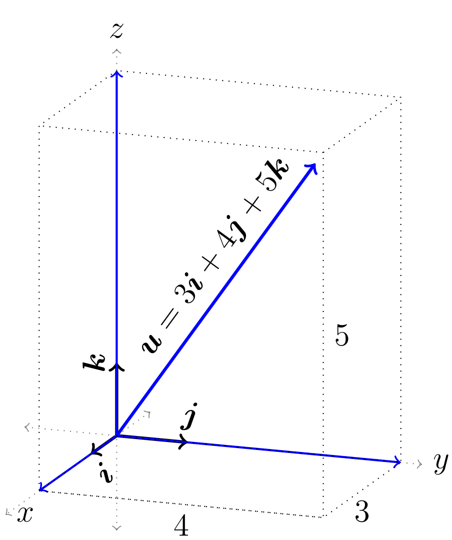
$$\boldsymbol{e}_1^3 = \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{e}_2^3 = \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{e}_3^3 = \boldsymbol{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Notation If it is clear that we are working in $\mathbb{R}^n$, then write $\boldsymbol{e}_i$ instead of $\boldsymbol{e}_i^n$. ♢

Every vector in $\mathbb{R}^n$ is a linear combination of the n standard basis vectors!

Example In $\mathbb{R}^3$:

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 3\boldsymbol{i} + 4\boldsymbol{j} + 5\boldsymbol{k} = 3\boldsymbol{e}_1^3 + 4\boldsymbol{e}_2^3 + 5\boldsymbol{e}_3^3$$



♢

Problem 12 Choose 2 vectors in each of $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$ and write them as linear combinations of standard basis vectors.

1.2.3 The "dot product" or "Euclidean inner product"

Given two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ both in $\mathbb{R}^n$, the *dot product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is given as

$$\boldsymbol{u} \boldsymbol{\cdot} \boldsymbol{v} \stackrel{\text{df}}{=} \sum_{i=1}^{n} u_i v_i$$

This is also called the *Euclidean inner product* or *standard real inner product* and is often denoted $\langle \boldsymbol{u} | \boldsymbol{v} \rangle$. We will adopt this notation and terminology when dealing with inner products of complex vectors as well as more general inner products.

The dot product is important and actually underlies both the notion of length of a vector and angle between vectors.

The *norm*[2] (*length*, *magnitude*) of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is defined by

$$\|\boldsymbol{v}\| \stackrel{\text{df}}{=} \sqrt{v_1^2 + \cdots + v_n^2} = \left(\sum_{i=1}^{n} v_i^2 \right)^{1/2}$$

the standard *Euclidean distance* is given by

$$\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) \stackrel{\text{df}}{=} \|\boldsymbol{u} - \boldsymbol{v}\| = \left(\sum_{i=1}^{n} (u_i - v_i)^2 \right)^{1/2}$$

[2]The 2-norm is just one of many possible norms on $\mathbb{R}^n$, for example the p-norm is $\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}$, where for $p = 1$ we have $\|\boldsymbol{v}\|_1 = \sum_{i=1}^{n} |v_i|$. We will almost exclusively use the 2-norm and hence drop the subscript 2, so $\|\boldsymbol{u}\| = \|\boldsymbol{u}\|_2$.

Problem 13 Write down two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in each of $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$ and compute $\|\boldsymbol{u}\|$, $\|3\boldsymbol{u} - 2\boldsymbol{v}\|$, and $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v})$.

Problem 14 Show that $\|\alpha\boldsymbol{u}\| = |\alpha|\|\boldsymbol{u}\|$.

From this we see

$$\left\|\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}\right\| = \frac{1}{\|\boldsymbol{u}\|}\|\boldsymbol{u}\| = 1$$

so

$$\hat{\boldsymbol{u}} = \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}$$

is the *unit vector* in the direction of $\boldsymbol{u}$.

Problem 15 (a) Find the unit vector in the direction of $(1, -2, 2)$.

(b) Find the vector of length 6 in the direction of $(2, -4, 2, 1)$.

(c) In $\mathbb{R}^2$, find the vector of length 4 which makes an angle of $\pi/4$ rad measured starting from the positive x-axis.

(d) In $\mathbb{R}^2$, find the vector of length $r > 0$ which makes an angle of θ rad measured starting from the positive x-axis.

Clearly,

$$\|\boldsymbol{u}\|^2 = \boldsymbol{u} \cdot \boldsymbol{u} = \sum_{i=1}^{n} u_i^2$$
$$\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v})^2 = \|\boldsymbol{v} - \boldsymbol{u}\|^2 = (\boldsymbol{v} - \boldsymbol{u}) \cdot (\boldsymbol{v} - \boldsymbol{u})$$

Problem 16 Show that the dot product satisfies each of the following

(a) (symmetry/commutativity) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$.

(b) (left linearity) $(\alpha \boldsymbol{u} + \beta \boldsymbol{v}) \cdot \boldsymbol{w} = \alpha(\boldsymbol{u} \cdot \boldsymbol{w}) + \beta(\boldsymbol{v} \cdot \boldsymbol{w})$.

(c) (positive definiteness) $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$ and equality holds iff $\boldsymbol{u} = \mathbf{0}$.

Problem 17 Show that

$$\mathrm{dist}(\boldsymbol{u}, \boldsymbol{v})^2 = \|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\boldsymbol{u} \cdot \boldsymbol{v}$$

The previous problem shows that

$$\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 \Leftrightarrow \boldsymbol{u} \cdot \boldsymbol{v} = 0$$

This is the **Pythagorean Theorem!** once we know (or define) that two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are *orthogonal* (perpendicular) exactly when $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$, then we can form the triangle with $\mathbf{0}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ as vertices. Let θ be the (unique) angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ such that $0 \le \theta \le \pi$, the law of cosines yields

$$\|\boldsymbol{v} - \boldsymbol{u}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos(\theta)$$

On the other hand Problem 17 gives

$$\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\boldsymbol{u} \cdot \boldsymbol{v}.$$

So we get

$$\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos(\theta) = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\boldsymbol{u} \cdot \boldsymbol{v}$$

and hence

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos(\theta).$$

Thus we have

$$\boxed{\cos(\theta) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}}$$

Problem 18 Characterize in terms of $\boldsymbol{u} \cdot \boldsymbol{v}$ the following:

(a) $\boldsymbol{u}$ and $\boldsymbol{v}$ are *orthogonal* (perpendicular).

(b) $\boldsymbol{u}$ and $\boldsymbol{v}$ are in the same direction.

(c) $\boldsymbol{u}$ and $\boldsymbol{v}$ are in opposite directions.

(d) $\boldsymbol{u}$ and $\boldsymbol{v}$ make an acute angle.

(e) $\boldsymbol{u}$ and $\boldsymbol{v}$ make an obtuse angle.

Problem 19 Show that in $\mathbb{R}^3$, $\boldsymbol{u} = \|\boldsymbol{u}\| \begin{bmatrix} \cos(\theta_1) \\ \cos(\theta_2) \\ \cos(\theta_3) \end{bmatrix}$ where θ_i is the angle between $\boldsymbol{u}$ and $\boldsymbol{e}_i$. (This generalizes to $\mathbb{R}^n$.) How does this compare to what happens in $\mathbb{R}^2$ where $\boldsymbol{u} = \|\boldsymbol{u}\| \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$?

Problem 20 Let

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

(a) Find a linear combination $\boldsymbol{v}$ of $\boldsymbol{u}_1, \boldsymbol{u}_2$ so that $\boldsymbol{v}$ and $\boldsymbol{u}_1$ are orthogonal. Geometrically, $\boldsymbol{v}$ is in the plane determined by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and is orthogonal to $\boldsymbol{u}_1$.

(b) Find $\boldsymbol{w}$ orthogonal to both $\boldsymbol{u}_1$ and $\boldsymbol{v}$.

1.3 (Complex) Euclidean vector spaces, $\mathbb{C}^n$

If the field of scalars is $\mathbb{C}$, then $\mathbb{R}^n$ is not closed under scalar multiplication, we must move to the vector spaces $\mathbb{C}^n$. Most everything about euclidean spaces caries over. The dot product is extended to the *standard inner product* on $\mathbb{C}^n$:[3]

$$\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \sum_{i=1}^{n} v_i^* u_i$$

Notice that $\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \boldsymbol{u} \cdot \boldsymbol{v}$ if $\boldsymbol{u}$ and $\boldsymbol{v}$ are in $\mathbb{R}^n$, so the inner product extends the dot product. [4]

The properties of dot products (see Problem 16) become:

Problem 21 Show that the inner product satisfies

(a) (conjugate symmetry) $\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \langle \boldsymbol{v} | \boldsymbol{u} \rangle^*$

[4] In order to not invalidate certain definitions that appear later, we will take $\boldsymbol{u} \cdot \boldsymbol{v}$ to be defined as before **even** when the vectors are complex. So for complex $\boldsymbol{u}$ and $\boldsymbol{v}$, typically, $\langle \boldsymbol{u} | \boldsymbol{v} \rangle \neq \boldsymbol{u} \cdot \boldsymbol{v}$.

(b) (left linearity) $\langle \alpha \boldsymbol{u} + \beta \boldsymbol{v} | \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{u} | \boldsymbol{w} \rangle + \beta \langle \boldsymbol{v} | \boldsymbol{w} \rangle$

(c) (positive definiteness) $\langle \boldsymbol{u} | \boldsymbol{u} \rangle \in \mathbb{R}^+ = [0, \infty)$ and $\langle \boldsymbol{u} | \boldsymbol{u} \rangle = 0$ iff $\boldsymbol{u} = \boldsymbol{0}$. (Notice $\langle \boldsymbol{u} | \boldsymbol{u} \rangle = \sum_{i=1}^{n} u_i^* u_i = \sum_{i=1}^{n} |u_i|^2$.)

Conjugate symmetry and left linearity give right *conjugate linearity*

$$\langle \boldsymbol{u} | \alpha \boldsymbol{v} + \beta \boldsymbol{w} \rangle = \alpha^* \langle \boldsymbol{u} | \boldsymbol{v} \rangle + \beta^* \langle \boldsymbol{u} | \boldsymbol{w} \rangle$$

From the positiveness of the inner product, it makes sense to define

$$\| \boldsymbol{u} \| = \sqrt{\langle \boldsymbol{u} | \boldsymbol{u} \rangle}$$

and it is simple to verify that this is a norm, i.e., the following are satisfied

- $\| \boldsymbol{u} \| = 0 \Rightarrow \boldsymbol{u} = \boldsymbol{0}$.
- $\| \boldsymbol{u} + \boldsymbol{v} \| \leq \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$. (*sub-additivity* or *triangle inequality*)
- $\| \alpha \boldsymbol{u} \| = |\alpha| \| \boldsymbol{u} \|$.

The triangle inequality requires Cauchy-Schwartz

$$|\langle \boldsymbol{u} | \boldsymbol{v} \rangle| \leq \| \boldsymbol{u} \| \, \| \boldsymbol{v} \|$$

Problem 22 Verify Cauchy-Schwartz by first verifying it for unit $\boldsymbol{u}$ and $\boldsymbol{v}$ and then generalizing. (Hint: Compute $\langle \boldsymbol{u} - \lambda \boldsymbol{v} | \boldsymbol{u} - \lambda \boldsymbol{v} \rangle$ and take $\lambda = \langle \boldsymbol{u} | \boldsymbol{v} \rangle$.) Next verify, the triangle inequality.

The next exercise indicates some things do change in passing to complex.

Problem 23 Show that

$$\langle \boldsymbol{u}|\boldsymbol{v}\rangle = 0 \Rightarrow \|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2,$$

but that

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 \not\Rightarrow \langle \boldsymbol{u}|\boldsymbol{v}\rangle = 0.$$

$\mathbb{C}$ can be interpreted/visualized as $\mathbb{R}^2$, i.e., $z = a + ib \mapsto (a, b)$. Get that every $z \in \mathbb{C}$ has the form $r(\cos(\theta) + i\sin(\theta))$ and notice $|z| = \|(a, b)\|$. This is indicated in the illustration on page 3.

1.4 Matrices

A $m \times n$ matrix is an array of scalars which is viewed as having m rows and n columns

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,j} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} & a_{i,2} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,j} & \cdots & a_{m,n} \end{bmatrix}$$

The i^{th} row will be denoted $\text{row}_i(A)$ and the j^{th} column will be denoted $\text{col}_j(A)$, where the A may be omitted if it is clear from the context.

$$\text{row}_i(A) = \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,j} & \cdots & a_{i,n} \end{bmatrix} \quad \text{col}_j(A) = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix}$$

Remark We will focus on a matrix as a collection of columns and/or rows more than an array of entries. This will be important in all that is to come! $\diamondsuit$

The i, j^{th} entry of A will be denoted $A_{i,j}$ and for an arbitrary matrix, A, we will often also use $A_{i,j} = a_{i,j}$.

$$\boxed{\textbf{The set of all } m \times n \textbf{ matrices is denoted } M_{mn}\textbf{.}}$$

Problem 24 Let $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & 0 \\ 4 & 1 & -2 \\ 2 & 5 & -6 \end{bmatrix}$

(a) What is $A_{1,2}$, $A_{4,2}$, and $A_{2,4}$?

(b) What is $\mathrm{row}_3(A)$, $\mathrm{col}_2(A)$?

Matrices will typically be denoted with uppercase letters, A, B, M, etc. The set of all $m \times n$ matrices will be denoted $M_{m,n}$.

We will encounter many special types of matrices as we go, a few initial types are:

- A is *square* iff A has the same number of rows and columns.
- A is *diagonal* iff $A_{i,j} = 0$ whenever $i \neq j$. (Note A need not be square!) For example

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \qquad \text{non square diagonal}$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{square diagonal}$$

- A square matrix A is *upper triangular* iff $A_{i,j} = 0$ whenever $i > j$, i.e., any non-zero entries occur in the upper right triangle. Similarly, A is *lower triangular* iff $A_{i,j} = 0$

whenever $j > i$. For example

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{upper triangular}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix} \qquad \text{lower triangular}$$

1.4.1 Linear combinations of matrices

Addition and scalar multiplication is very much like that for vectors: Let A and B be two $m \times n$ matrices:

- (vector addition/component-wise addition)

$$(A + B)_{i,j} = A_{i,j} + B_{i,j}$$

- (scalar multiplication/component-wise multiplication)

$$(\alpha A)_{i,j} = \alpha A_{i,j}$$

For example if $A = \begin{bmatrix} 1 & 2 \\ -3 & 1 \\ 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 3 \\ -4 & 0 \\ 1 & 1 \end{bmatrix}$, then

$$A + B = \begin{bmatrix} 1 & 5 \\ -7 & 1 \\ 3 & 2 \end{bmatrix} \qquad -3A = \begin{bmatrix} -3 & -6 \\ 9 & -3 \\ -6 & -3 \end{bmatrix}$$

Problem 25 For the A and B in the above example find $2A - 3B$.

1.4.2 Adjoint and Transpose

Given an $m \times n$ matrix A, the *transpose* of A is the $n \times m$ matrix A^T where the rows/columns of A are now the columns/rows of A^T, formally this amounts to

$$A^T_{i,j} = A_{j,i}$$

For example:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 6 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 6 \end{bmatrix}$$

The following *fact* would really be a better definition for A^T:

$$\mathrm{row}_i(A^T) = \mathrm{col}_i(A) \qquad \mathrm{col}_i(A^T) = \mathrm{row}_i(A)$$

This seemingly simple operation will turn out to be of tremendous importance.

Problem 26 (a) Show that $(\alpha A)^T = \alpha(A^T)$ and $(A+B)^T = A^T + B^T$. So the transpose operation is a linear operator from M_{mn} to M_{nm}, i.e., $(\alpha A + \beta B)^T = \alpha A^T + \beta B^T$.

(b) Show $(A^T)^T = A$

Among the special kinds of matrices we will use add the following:

- A is *symmetric* iff $A^T = A$.
- A is *skew-symmetric* iff $A^T = -A$.

Problem 27 Show that every square matrix, A, can be written as $A = S + T$ where S is symmetric and T is skew symmetric.

Hint: Consider $A + A^T$ and $A - A^T$.

Most matrices we consider will have real entries, however, there are important results that require considering matrices with complex entries.

If A is a matrix with complex entries, then the *adjoint* or *conjugate transpose* of A is A^* and is defined by

$$A^*_{i,j} = (A_{j,i})^*$$

For example

$$\begin{bmatrix} 1+i & 2 & 3-2i \\ 0 & 1-2i & 6i \end{bmatrix}^* = \begin{bmatrix} 1-i & 0 \\ 2 & 1+2i \\ 3+2i & -6i \end{bmatrix}.$$

Note that for A a complex matrix we have two distinct operations A^T and A^*, these both extend the transpose operation on real matrices.

Problem 28 Show that $(\alpha A)^* = \alpha^* A^*$.

A is *Hermitian* iff $A^* = A$. For A with real entries $A^* = A^T$, in which case Hermitian and symmetric mean the same thing.

1.4.3 Matrix Multiplication

We will consider 5 ways to multiply matrices, all of which give the same result and each of which is useful and will be used.

Row×Column: This is the "standard way" matrix multiplication is introduced. For A is an $m \times n$ matrix and B is an $n \times k$ matrix, then we can form the product $AB \in m \times k$ defined by

$$(AB)_{il} = \text{row}_i(A) \cdot \text{col}_l(B) = \sum_{j=1}^{n} A_{i,j} B_{j,l}$$

$$\begin{bmatrix} \text{row}_1(A) \\ \hdashline \text{row}_2(A) \\ \hdashline \vdots \\ \hdashline \text{row}_i(A) \\ \hdashline \vdots \\ \hdashline \text{row}_m(A) \end{bmatrix} \left[\begin{array}{c:c:c:c:c:c} \text{col}_1(B) & \text{col}_2(B) & \cdots & \text{col}_j(B) & \cdots & \text{col}_k(B) \end{array}\right] =$$

$$\begin{bmatrix} \text{row}_1(A) \cdot \text{col}_1(B) & \cdots & \text{row}_1(A) \cdot \text{col}_j(B) & \cdots & \text{row}_1(A) \cdot \text{col}_k(B) \\ \text{row}_2(A) \cdot \text{col}_1(B) & \cdots & \text{row}_2(A) \cdot \text{col}_j(B) & \cdots & \text{row}_2(A) \cdot \text{col}_k(B) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \text{row}_m(A) \cdot \text{col}_1(B) & \cdots & \text{row}_m(A) \cdot \text{col}_j(B) & \cdots & \text{row}_m(A) \cdot \text{col}_k(B) \end{bmatrix}$$

Example

$$\begin{bmatrix} 1 & 3 \\ \hdashline 2 & 4 \\ \hdashline -1 & -1 \end{bmatrix} \left[\begin{array}{c:c:c} 1 & 0 & 4 \\ 1 & -2 & 6 \end{array}\right]$$

$$= \begin{bmatrix} (1,3) \cdot (1,1) & (1,3) \cdot (0,-2) & (1,3) \cdot (4,6) \\ (2,4) \cdot (1,1) & (2,4) \cdot (0,-2) & (2,4) \cdot (4,6) \\ (-1,-1) \cdot (1,1) & (-1,-1) \cdot (0,-2) & (-1,-1) \cdot (4,6) \end{bmatrix}$$

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

$$= \begin{bmatrix} 4 & -6 & 22 \\ 6 & -8 & 32 \\ -2 & 2 & -10 \end{bmatrix}$$

$\diamondsuit$

Three important facts about matrix multiplication are:

- Matrix multiplication is **not commutative**, that is, in general, $AB = BA$ fails. For this to be defined A and B must be square.

- Matrix multiplication is linear,

$$A(\alpha B + \beta C) = \alpha(AB) + \beta(AC)$$
$$(\alpha B + \beta C)A = \alpha(BA) + \beta(CA).$$

- Matrix multiplication **is associative**, that is,

$$(AB)C = A(BC).$$

The proof of linearity and non-commutativity is an exercise. The proof of associativity can proceed either by brute force (unenlightening) or by a connection we will later develop between matrices and linear functions (then associativity becomes trivial). Part of this relationship is already apparent. If $A \in M_{mn}$, then multiplication on the right by A defines a mapping, which we call L_A from $\mathbb{R}^n$ to $\mathbb{R}^m$, namely, $L_A(\boldsymbol{x}) = A\boldsymbol{x}$ which is linear

$$L_A(\underbrace{\alpha \boldsymbol{u} + \beta \boldsymbol{v}}_{\text{linear combination of } \boldsymbol{u} \text{ and } \boldsymbol{v}}) = \underbrace{\alpha L_A(\boldsymbol{u}) + \beta L_A(\boldsymbol{v})}_{\text{the same linear combination of } L_A(\boldsymbol{u}) \text{ and } L_A(\boldsymbol{v})}$$

Problem 29 (a) Produce 2×2 matrices A and B so that $AB \neq BA$.

(b) Show by example that *cancellation* fails, i.e., find A, B, C so that $AB = AC$ and yet $B \neq C$.

Problem 30 Find at least three (perhaps infinitely many) "square roots of" $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Problem 31 Show that $(AB)^T = B^T A^T$. Similarly, $(AB)^* = B^* A^*$ for complex A, B.

Problem 32 (a) Show that AA^T and $A^T A$ are symmetric, and that AA^* and $A^* A$ are Hermitian (for complex A).

(b) Show, in addition, that $\boldsymbol{x}^T (A^T A)\boldsymbol{x} \geq 0$. (Similarly for the other 3 variants AA^T, $A^* A$, and AA^*.)[5]

Dot product/inner product as matrix multiplication. We defined matrix multiplication via the dot product, this can be

[5] This shows that $A^T A$ is *positive semidefinite*. If we wanted additionally to get that $\boldsymbol{x}^* (A^* A)\boldsymbol{x} = 0$ iff $\boldsymbol{x} = \boldsymbol{0}$ (*positive definite*), then we would require that $A^* A$ be invertible, we will see that this occurs precisely when A has maximal rank.

turned around:

Problem 33 Show that $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v}^T \boldsymbol{u}$, while for $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{C}^n$, $\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \boldsymbol{v}^* \boldsymbol{u}$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^n$, when $\boldsymbol{u}$ and $\boldsymbol{v}$ are viewed as column vectors. Thus both the dot product and inner product can be defined by matrix multiplication.

Columns×Rows: The preceding problem begs us to define the *outer product*

Problem 34 Define the outer product of $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^n$ by

$$\boldsymbol{u} \otimes \boldsymbol{v} \stackrel{\text{df}}{=} \boldsymbol{u}\boldsymbol{v}^T \in M_{mn}$$

Compute $(2,3) \otimes (1,-1,3)$ and $(1,-1,3) \otimes (2,3)$. What is their relationship? What is the relationship between $\boldsymbol{u} \otimes \boldsymbol{v}$ and $\boldsymbol{v} \otimes \boldsymbol{u}$ in general?

Notice for $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^n$

$$\boldsymbol{u} \otimes \boldsymbol{v} = \boldsymbol{u}\boldsymbol{v}^T = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \left[\, \boldsymbol{u}v_1 \,\vdots\, \boldsymbol{u}v_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{u}v_n \,\right] = \begin{bmatrix} u_1\boldsymbol{v} \\ \hdashline u_2\boldsymbol{v} \\ \hdashline \vdots \\ \hdashline u_m\boldsymbol{v} \\ \hdashline \end{bmatrix}$$

So all columns/rows are co-linear, later we will call such a matrix rank 1 and all rank 1 matrices in M_{mn} arise as $\boldsymbol{u}\boldsymbol{v}^T$ for some $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^n$.

Matrix multiplication could be defined with outer product. For A a $m \times n$ matrix and B a $n \times k$ matrix we have

$$AB = \left[\, \mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A) \,\right] \begin{bmatrix} \mathrm{row}_1(B) \\ \hdashline \mathrm{row}_2(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_i(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_n(B) \end{bmatrix}$$

$$= \sum_{l=1}^{k} \mathrm{col}_l(A) \otimes \mathrm{row}_l(B)$$

Here I am treating $\mathrm{row}_l(B)$ as an element of R^k and hence as a column vector, if we literally take it as a row vector, then we should simply write $AB = \sum_{l=1}^{n} \mathrm{col}_l(A)\,\mathrm{row}_l(B)$.

$$\begin{aligned} \left(\sum_{l=1}^{n} \mathrm{col}_l(A) \otimes \mathrm{row}_l(B) \right)_{i,j} &= \left(\sum_{l=1}^{n} \mathrm{col}_l(A) \otimes \mathrm{row}_l(B) \right)_{i.j} \\ &= \sum_{l=1}^{n} \left(\mathrm{col}_l(A) \otimes \mathrm{row}_l(B) \right)_{i,j} \\ &= \sum_{l=1}^{n} \mathrm{col}_l(A)_i \, \mathrm{row}_l(B)_j \\ &= \sum_{l=1}^{n} A_{i,l} B_{l,j} = (AB)_{i,j} \end{aligned}$$

Example

$$\left[\begin{array}{c:c} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{array}\right] \left[\begin{array}{ccc} 1 & 0 & 4 \\ \hdashline 1 & -2 & 6 \end{array}\right] = (1,2,-1) \otimes (1,0,4) + (3,4,-1) \otimes (1,-2,6)$$
$$= \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -2 \\ 6 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 6 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 4 \\ 2 & 0 & 8 \\ -1 & 0 & -4 \end{bmatrix} + \begin{bmatrix} 3 & -6 & 18 \\ 4 & -8 & 24 \\ -1 & 2 & -6 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & -6 & 22 \\ 6 & -8 & 32 \\ -2 & 2 & -10 \end{bmatrix}$$

$\diamondsuit$

Columns×Columns: Here we describe matrix multiplication AB by viewing the columns of the result as linear combinations of the columns of A.

For A an $m \times n$ matrix view A alternatively as a matrix of columns

$$A = \left[\, \mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A) \,\right]$$

where $\mathrm{col}_j(A) \in \mathbb{R}^m$.

Right multiplication of A by a vector $\boldsymbol{x}$ in $\mathbb{R}^n$ becomes a linear combination of columns:

$$A\boldsymbol{x} = \left[\, \mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A) \,\right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$
$$= x_1 \mathrm{col}_1(A) + \cdots + x_n \mathrm{col}_n(A)$$

Example

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} = (4) \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + (6) \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} 22 \\ 32 \\ -10 \end{bmatrix}$$

$\diamondsuit$

The Matrix product AB becomes

$$AB = A \left[\, \mathrm{col}_1(B) \,\vdots\, \mathrm{col}_2(B) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(B) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_k(B) \,\right]$$
$$= \left[\, A\,\mathrm{col}_1(B) \,\vdots\, A\,\mathrm{col}_2(B) \,\vdots\, \cdots \,\vdots\, A\,\mathrm{col}_j(B) \,\vdots\, \cdots \,\vdots\, A\,\mathrm{col}_k(B) \,\right]$$

where

$$A\,\mathrm{col}_l(B) = \left[\mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A)\right] \mathrm{col}_l(B)$$

$$= \left[\mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_j(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A)\right] \begin{bmatrix} B_{1,l} \\ B_{2,l} \\ \vdots \\ B_{n,l} \end{bmatrix}$$

Example

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{bmatrix} \left[\begin{array}{c:c:c} 1 & 0 & 4 \\ 1 & -2 & 6 \end{array}\right] =$$

$$= \left[\begin{array}{c:c:c} \left[\begin{array}{c:c} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{array}\right] \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \left[\begin{array}{c:c} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{array}\right] \begin{bmatrix} 0 \\ -2 \end{bmatrix} & \left[\begin{array}{c:c} 1 & 3 \\ 2 & 4 \\ -1 & -1 \end{array}\right] \begin{bmatrix} 4 \\ 6 \end{bmatrix} \end{array}\right]$$

$$= \left[\begin{array}{c:c:c} (1)\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + (1)\begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} & (0)\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + (-2)\begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} & (4)\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + (6)\begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} \end{array}\right]$$

$$= \begin{bmatrix} 4 & -6 & 22 \\ 6 & -8 & 32 \\ -2 & 2 & -10 \end{bmatrix}$$

$\diamondsuit$

Rows×Rows: Here we describe matrix multiplication AB by viewing the columns of the result as linear combinations of the columns of A.

Let B be an $n \times k$ matrix, let $\boldsymbol{x} \in \mathbb{R}^n$, typically we view $\boldsymbol{x}$ as a column vector and need to write $\boldsymbol{x}^T B$ to make sense of the multiplication, when this is clear (and unambiguous), we will be sloppy and just write $\boldsymbol{x}B$. (So we have a uniform notion of left and right multiplication of a matrix by a vector.)

We have

$$\boldsymbol{x}B = \boldsymbol{x} \left[\begin{array}{c} \mathrm{row}_1(B) \\ \hdashline \mathrm{row}_2(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_i(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_n(B) \end{array}\right]$$

$$= x_1\,\mathrm{row}_1(B) + \cdots + x_n\,\mathrm{row}_n(B) \in R^k.$$

so the result is a linear combination of rows of B.

If A is an $m \times n$ matrix also viewed as rows, then

$$AB = \begin{bmatrix} \mathrm{row}_1(A) \\ \hdashline \mathrm{row}_2(A) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_i(A) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_m(A) \end{bmatrix} B = \begin{bmatrix} \mathrm{row}_1(A)\, B \\ \hdashline \mathrm{row}_2(A)\, B \\ \hdashline \vdots \\ \hdashline \mathrm{row}_i(A)\, B \\ \hdashline \vdots \\ \hdashline \mathrm{row}_m(A)\, B \end{bmatrix}$$

where in turn

$$\begin{aligned} \mathrm{row}_i(A)\, B &= \begin{bmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,n} \end{bmatrix} \begin{bmatrix} \mathrm{row}_1(B) \\ \hdashline \mathrm{row}_2(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_i(B) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_n(B) \end{bmatrix} \\ &= A_{i,1}\,\mathrm{row}_1(B) + \cdots + A_{i,n}\,\mathrm{row}_n(B) \in R^k \end{aligned}$$

Example

$$\begin{aligned} \begin{bmatrix} 1 & 3 \\ \hdashline 2 & 4 \\ \hdashline -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ 1 & -2 & 6 \end{bmatrix} &= \begin{bmatrix} \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ \hdashline 1 & -2 & 6 \end{bmatrix} \\ \hdashline \begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ \hdashline 1 & -2 & 6 \end{bmatrix} \\ \hdashline \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ \hdashline 1 & -2 & 6 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} (1) \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} + (3) \begin{bmatrix} 1 & -2 & 6 \end{bmatrix} \\ \hdashline (2) \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} + (4) \begin{bmatrix} 1 & -2 & 6 \end{bmatrix} \\ \hdashline (-1) \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} + (-1) \begin{bmatrix} 1 & -2 & 6 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 4 & -6 & 22 \\ 6 & -8 & 32 \\ -2 & 2 & -10 \end{bmatrix} \end{aligned}$$

$\diamondsuit$

Problem 35 Show that $A\boldsymbol{e}_i^n = \mathrm{col}_i(A)$ and $\boldsymbol{e}_i^m A = \mathrm{row}_i(A)$

Notice that this problem tells you how to find a matrix A given that you know how the matrix acts on the standard basis vectors.

Problem 36 The function R_θ which rotates the plane counterclockwise through θ radians is given by a matrix, that is,

$$R_\theta\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = A\begin{bmatrix} x \\ y \end{bmatrix}$$

for some $\square \times \square$ matrix A. find the matrix by computing $A\boldsymbol{e}_1^2$ and $A\boldsymbol{e}_2^2$ and then using Problem 35.

Problem 37 The function that reflects points in $\mathbb{R}^2$ through the line $2x - 3y = 0$ is given by a $\boxed{} \times \boxed{}$ matrix A. Find A.

The following also follows from Problem 35

Problem 38 We saw before that "cancellation" fails for matrices (see Problem 29). However there is an important sense in which it holds:

(a) Show that if A and B are $m \times n$ matrix such that $A\boldsymbol{x} = B\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^n$, then $A = B$.

(b) Show that if A and B are $m \times n$ matrix such that $\boldsymbol{x}^T A = \boldsymbol{x}^T B$ for all $\boldsymbol{x} \in \mathbb{R}^m$, then $A = B$.

Define $I_k \in M_{kk}$ by

$$(I_k)_{i,j} = \begin{bmatrix} \boldsymbol{e}_1^k & \boldsymbol{e}_2^k & \cdots & \boldsymbol{e}_k^k \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1^k \\ \boldsymbol{e}_2^k \\ \cdots \\ \boldsymbol{e}_k^k \end{bmatrix} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Then for $A \in M_{mn}$

$$AI_n = I_m A = A$$

To see this notice

$$I_m A = \begin{bmatrix} \boldsymbol{e}_1^m \\ \hdashline \boldsymbol{e}_2^m \\ \hdashline \vdots \\ \hdashline \boldsymbol{e}_m^m \end{bmatrix} A = \begin{bmatrix} \boldsymbol{e}_1^m A \\ \hdashline \boldsymbol{e}_2^m A \\ \hdashline \vdots \\ \hdashline \boldsymbol{e}_m^m A \end{bmatrix} = \begin{bmatrix} \mathrm{row}_1(A) \\ \hdashline \mathrm{row}_2(A) \\ \hdashline \vdots \\ \hdashline \mathrm{row}_m(A) \end{bmatrix} = A$$

similarly

$$\begin{aligned} AI_n &= A \left[\boldsymbol{e}_1^n \,\vdots\, \boldsymbol{e}_2^n \,\vdots\, \cdots \,\vdots\, \boldsymbol{e}_n^n \right] = \left[A\boldsymbol{e}_1^n \,\vdots\, A\boldsymbol{e}_2^n \,\vdots\, \cdots \,\vdots\, A\boldsymbol{e}_n^n \right] \\ &= \left[\mathrm{col}_1(A) \,\vdots\, \mathrm{col}_2(A) \,\vdots\, \cdots \,\vdots\, \mathrm{col}_n(A) \right] = A \end{aligned}$$

Problem 39 For A a 4×3 matrix

(a) What matrix would we multiply A by on the left so that the result is what we get by replacing $\mathrm{row}_2(A)$ by $\mathrm{row}_2(A) + \alpha\,\mathrm{row}_3(A)$?

(b) What matrix would we multiply on the left by to interchange the 1^{st} and 4^{th} rows of A?

(c) What matrix would we multiply on the right to replace $\mathrm{col}_3(A)$ by $\mathrm{col}_3(A) + \alpha\mathrm{col}_1(A)$?

Problem 40 For A and $m \times n$ matrix show that $A_{ij} = (A\boldsymbol{e}_j^n) \cdot \boldsymbol{e}_i^m$. What is $\boldsymbol{e}_i^m \cdot (A\boldsymbol{e}_j^n)$?

Problem 41 (a) For A a real matrix, show that A^T is the unique real matrix, B, so that

$$(A\boldsymbol{x}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot (B\boldsymbol{y})$$

Hint: Use Problem 40 or Problem 38.

(b) For A a complex matrix, show that A^* is the unique complex matrix, B, so that

$$\langle A\boldsymbol{x}|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|B\boldsymbol{y}\rangle^{6}$$

[6] This is will become the definition of adjoint in the context of inner product spaces.

Block Matrices The matrices A and B may be partitioned into arbitrary blocks

$$A = \left[\begin{array}{c|c|c} A_1 & A_2 & A_3 \\ \hline A_4 & A_5 & A_6 \end{array}\right]$$

$$B = \left[\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \\ \hline B_5 & B_6 \end{array}\right]$$

$$AB = \left[\begin{array}{c|c} A_1B_1 + A_2B_3 + A_3B_5 & A_1B_2 + A_2B_4 + A_3B_6 \\ \hline A_4B_1 + A_5B_3 + A_6B_5 & A_4B_2 + A_5B_4 + A_6B_6 \end{array}\right]$$

Of course all of the sizes have to match up so that it makes sense to multiply the blocks.

Example

$$\left[\begin{array}{cc} 1 & 3 \\ 2 & 4 \\ \hline -1 & -1 \end{array}\right] \left[\begin{array}{cc|c} 1 & 0 & 4 \\ 1 & -2 & 6 \end{array}\right] = \left[\begin{array}{c|c} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -2 \end{bmatrix} & \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} \\ \hline \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -2 \end{bmatrix} & \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \begin{bmatrix} 4 & -6 \\ 6 & -8 \end{bmatrix} & \begin{bmatrix} 22 \\ 32 \end{bmatrix} \\ \hline \begin{bmatrix} -2 & 2 \end{bmatrix} & -10 \end{array}\right] = \begin{bmatrix} 4 & -6 & 22 \\ 6 & -8 & 32 \\ -2 & 2 & -10 \end{bmatrix}$$

Note that this has the "block form"

$$\left[\begin{array}{c} A_1 \\ \hdashline A_2 \end{array}\right] \left[\begin{array}{c;{2pt/2pt}c} B_1 & B_2 \end{array}\right] = \left[\begin{array}{c;{2pt/2pt}c} A_1B_1 & A_1B_2 \\ \hdashline A_2B_1 & A_2B_2 \end{array}\right]$$

◊

Problem 42 For each of the following matrices A and vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ write $A\boldsymbol{x}$ as a linear combination of the columns of A and $\boldsymbol{y}A$ as a linear combination of the rows of A.

(a) $A = \begin{bmatrix} 1 & 0 & 3 \\ 4 & -1 & 2 \end{bmatrix}$, $\boldsymbol{x} = (1, 2, -2)$, and $\boldsymbol{y} = (4, -1)$.

(b) $A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \\ 5 & 2 \end{bmatrix}$, $\boldsymbol{x} = (2, -2)$, and $\boldsymbol{y} = (4, -1, 3)$.

Problem 43 Expand each of the following matrix multiplications AB in each of "rows×columns", "columns×rows", "columns×columns", and "rows×rows":

(a) $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ 0 & 4 \\ -3 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 2 & -3 \\ 4 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 & 1 \\ -2 & 0 & 2 & 1 \end{bmatrix}$

Problem 44 Compute AB using "block multiplication"

$$A = \left[\begin{array}{ccc:cc} 1 & 0 & -2 & 3 & 4 \\ 0 & 2 & 1 & 3 & -3 \\ \hdashline 5 & 2 & -2 & 1 & 0 \end{array}\right] \qquad B = \left[\begin{array}{cc:c} 3 & -1 & 6 \\ 1 & 2 & -2 \\ -3 & 2 & 0 \\ \hdashline 0 & 2 & 2 \\ 1 & -1 & -2 \end{array}\right]$$

Problem 45 A square $n \times n$ matrix, A, is *positive definite* iff $\boldsymbol{x}^T A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$. Let B be the upper left $k \times k$ square submatrix of A. Show that if A is positive definite so is B.

Problem 46 Describe the result of multiplying two square diagonal matrices, lower triangular matrices, upper triangular matrices. What are the diagonal entries?

Problem 47 Let $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Compute, by hand, A^5.

2 $Ax = b$

Consider the problem of solving

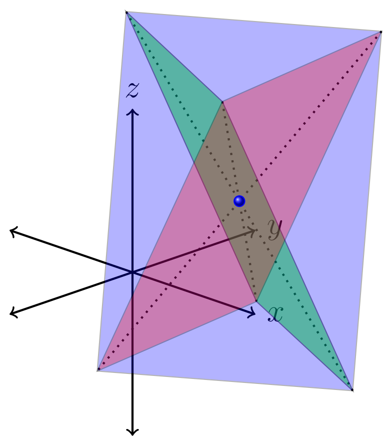
$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + x_3 &= 1 \\ 2x_1 + 3x_2 - 4x_3 &= 1 \end{aligned}$$

The matrix form of this is

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 3 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

There are two pictures associated to this problem:

The row picture: Here we view this as asking for a point where the three planes described by the three equations intersect.

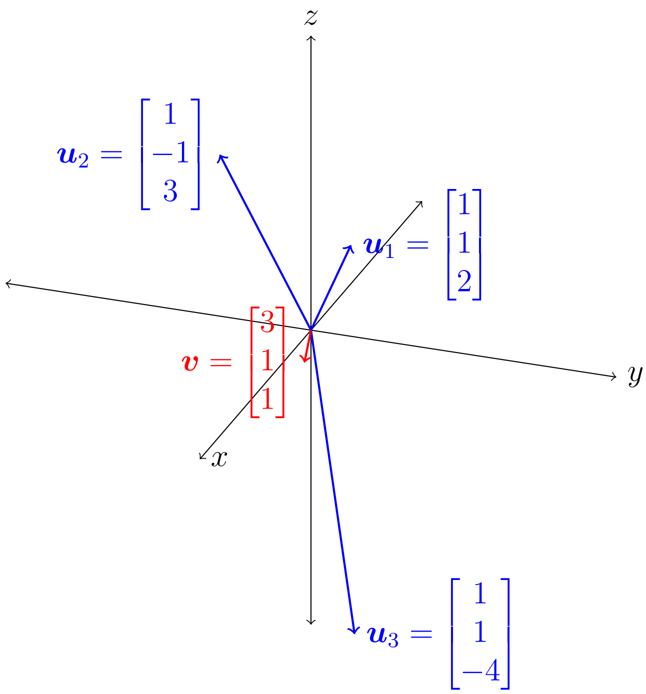


This illustrates the three intersecting planes whose equations are given.

The column picture: Here we view this equation as a linear combination of columns.

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

It is clear that $x_1 = x_2 = x_3 = 1$ will work.



This picture illustrates the three vectors for which we want to find a linear combination of giving the vector $(1, 3, 1)$.

Problem 48 Describe geometrically, the set of all linear combinations of $(1, 1, 2)$, $(1, -1, 3)$, and $(1, 1, -4)$?

Problem 49 Illustrate the row picture and column picture for the solution to

$$\begin{aligned} 2x_1 + 3x_2 &= 1 \\ x_1 + x_2 &= 0 \end{aligned}$$

Note that the row picture will consist of two intersecting lines.

If you had just one equation to solve

$$ax = b,$$

then of course you would simply multiply both sides by $a^{-1} = \frac{1}{a}$. In the case that there are n-equations in n-unknowns, so that in the equation

$$A\boldsymbol{x} = \boldsymbol{b}$$

we will look for an analogous multiplicative inverse A^{-1} for the square $n \times n$ matrix A.

Important! For a square matrix A, A^{-1} will be the unique matrix B so that $AB = BA = I_n$. A big question will be "*When does a square matrix A have an inverse?*". A square matrix is called *singular* if it fails to have an inverse and *non-singular* if invertible.

Problem 50 Show that if A is invertible, then A^{-1} is unique.

We will see later (see Problem 63) that for square matrices A the following are equivalent:

- There is a left inverse, i.e., a B so that $BA = I$.
- There is a right inverse, i.e., a C so that $AC = I$.
- A is invertible.

Moreover, in the first two cases, $B = C = A^{-1}$.

This will follow from the fact that the function $L_A : \boldsymbol{x} \mapsto A\boldsymbol{x}$ the following are equivalent

- L_A is is one-to-one.
- L_A is onto.
- L_A is invertible.

Problem 51 Show that is A is invertible, then so is A^k. What is $(A^k)^{-1}$? Define $A^{-k} = (A^k)^{-1}$, what is A^{n+m} for arbitrary integers n, m?

Problem 52 Show that if A and B are invertible, then so is AB and $(AB)^{-1} = B^{-1}A^{-1}$.[8]

Problem 53 Show that if A is invertible, then so is A^* and $(A^*)^{-1} = (A^{-1})^*$.

Problem 54 Characterize when a diagonal matrix A is invertible. For A that is diagonal and invertible, compute A^{-1}. (For an arbitrary $n \times n$ diagonal matrix A!)

[8] We will see later that AB is invertible iff A and B are invertible, assuming all are square.

The next problem indicates how to invert certain block diagonal matrices.

Problem 55 (a) Suppose $B \in M_{nn}$ and $D \in M_{mm}$ are square invertible matrices, $C, X \in M_{nm}$, and O is the $m \times n$ matrix of 0's. Compute

$$\left[\begin{array}{c:c} B & C \\ \hdashline O & D \end{array}\right] \left[\begin{array}{c:c} B^{-1} & X \\ \hdashline O & D^{-1} \end{array}\right]$$

(b) Use (a) to find X (in terms of B, B^{-1}, D, D^{-1}, and C) so that

$$\left[\begin{array}{c:c} B & C \\ \hdashline O & D \end{array}\right] \left[\begin{array}{c:c} B^{-1} & X \\ \hdashline O & D^{-1} \end{array}\right] = I_{n+m}.$$

This shows that $A = \left[\begin{array}{c:c} B & C \\ \hdashline O & D \end{array}\right]$ is invertible iff B and D are invertible.

Problem 56 Use the preceding problem to show that a triangular matrix is invertible iff there are **no** 0's on the diagonal.

Problem 57 Use the preceding problem to find A^{-1} for

$$A = \begin{bmatrix} 1 & 0 & 0 & -1 & 2 \\ 0 & 2 & 0 & 3 & 1 \\ 0 & 0 & -3 & -2 & 3 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix}$$

2.1 Gauss elimination

Gauss elimination is a key, and simple, technique underlying many procedures and results.

Consider again the system of 3 equations and 3 unknowns:

$$\begin{aligned} x_1 + \ x_2 + \ x_3 &= 3 && \text{(eq 1)} \\ x_1 - \ x_2 + \ x_3 &= 1 && \text{(eq 2)} \\ 2x_1 + 3x_2 - 4x_3 &= 1 && \text{(eq 3)} \end{aligned}$$

To solve this we may first *eliminate* the x_1's from the second two equations by replacing (eq 2) with $-$(eq 1) + (eq 2) and replacing (eq 3) similarly with -2(eq 1) + (eq 3) this results in a new system of equations, which clearly has the same solutions as the original system:

$$\begin{aligned} x_1 + \ x_2 + \ x_3 &= 3 && \text{(eq 1)} \\ -2x_2 \qquad &= -2 && \text{(eq 2)} \\ x_2 - 6x_3 &= -5 && \text{(eq 3)} \end{aligned}$$

Next we can eliminate the x_2 in (eq 3) by replacing (eq 3) with $1/2$(eq 2) + (eq 3):

$$\begin{aligned} x_1 + \ x_2 + \ x_3 &= 3 && \text{(eq 1)} \\ -2x_2 \qquad &= -2 && \text{(eq 2)} \\ -6x_3 &= -6 && \text{(eq 3)} \end{aligned}$$

This can now easily be solved by *back substitution*

$$\begin{aligned} x_3 &= 1 \\ x_2 &= 1 \\ x_1 &= 3 - 1 - 1 = 1 \end{aligned}$$

So the unique solution is $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Two systems of m linear equations in n unknowns are called *equivalent* iff they both have the same solution set. There are three elementary operations that transform systems of equations into equivalent systems of equations:

(1) Replace the (eq i) by (eq i)+α(eq j) for some scalar α.

(2) Swap (eq i) and (eq j).

(3) Multiply (eq i) by a non-zero scalar.

We shall see, these operations suffice to solve any system. First we translate these operations into operations on matrices.

Reconsider what we just did in the language of matrices. Start with the *augmented matrix*, where A is the coefficient matrix and $\boldsymbol{b} = (3, 1, 1)$:

$$\left[A \,\vdots\, \boldsymbol{b} \right] = \left[\begin{array}{ccc:c} \boxed{1} & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ 2 & 3 & -4 & 1 \end{array}\right]$$

The $(1,1)$ entry is called the *pivot* (a pivot must always be non-zero.)

Our first step corresponds to replacing row 2 by the result of subtracting row 1 from row 2, denoted $\text{row}_2 \Leftarrow \text{row}_2 - \text{row}_1$. The point is to *eliminate* the $(2,1)$ entry. The result is

$$\left[\begin{array}{ccc:c} \boxed{1} & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 2 & 3 & -4 & 1 \end{array}\right] \qquad (2,1) \text{ elimination}$$

Next we eliminate the $(3,1)$ entry in a similar fashion giving

$$\left[\begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 0 & \boxed{-2} & 0 & -2 \\ 0 & 1 & -6 & -5 \end{array}\right] \qquad (3,1) \text{ elimination}$$

Now we eliminate the $(3,2)$ element using the new pivot -2 by replacing row 3 by the result of multiplying row 2 by $1/2$ and adding this to row 3. The result is

$$\left[\begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & \boxed{-6} & -6 \end{array}\right] \qquad (3,2) \text{ elimination}$$

This results in the final pivot -6.

Call the **upper triangular matrix** that results from applying this procedure to just the coefficient matrix A, U, that is,

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{bmatrix}$$

In general the result will not be triangular, for example, the number of equations and unknowns might not be the same.

A matrix A is in *echelon form* if the first non-zero entry in a particular row occurs to the right of the first non-zero entry from the preceding rows. Echelon form is a generalization of upper triangular.

For example

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

is in echelon form while

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

is not, although it is clear that one row operation would transform this into echelon form.

The elementary operations performed on systems of equations above give rise to *elementary row operations* on matrices:

The three *types* of elementary row operations.

Type I: Replace row_i with $\text{row}_i + \alpha\,\text{row}_j$.

Type II: Swap row_i and row_j.

Type III: Replace row_i by $\alpha\,\text{row}_i$ for $\alpha \neq 0$.

It is evident that every matrix is row equivalent to a matrix in echelon form. In fact it is evident that this is true even if we drop Type III operations. **Make sure you convince yourself of this!**

Generally move the left most non-zero entry to the first row using a Type II operation, then kill all non-zero entries below it, this will be a sequence of Type I operations, find the next left most non-zero entry in a row below the first one, move it to the second row, kill all the non-zero entries below it, etc.

Example 1 Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 6 & 12 & 0 & 6 & -6 \\ 3 & 6 & 1 & 7 & 0 \\ 2 & 4 & 0 & 2 & -2 \end{bmatrix}$$

Use the $(1,1)$ entry to kill all non-zero entries below using Type I operations $\text{row}_2 \Leftarrow \text{row}_2 - 6\,\text{row}_1$, $\text{row}_3 \Leftarrow \text{row}_3 - 3\,\text{row}_1$, and $\text{row}_4 \Leftarrow \text{row}_4 - 2\,\text{row}_1$. These are the $(2,1)$, $(3,1)$, and $(4,1)$ steps in the elimination process. The result is

$$\begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 0 & 0 & 6 & 24 & 6 \\ 0 & 0 & 4 & 16 & 6 \\ 0 & 0 & 2 & 8 & 2 \end{bmatrix}$$

Now the first non-zero entry in a later row occurs in the 3$^{\text{rd}}$ column. So start eliminating in the 3$^{\text{rd}}$ column, again using Type I operations, $\text{row}_3 \Leftarrow \text{row}_3 - \frac{4}{6}\,\text{row}_2$, and $\text{row}_4 \Leftarrow \text{row}_4 - \frac{2}{6}\,\text{row}_2$. These are the $(3,3)$ and $(4,3)$ steps of the elimination. The result is

$$\begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 0 & 0 & 6 & 24 & 6 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This is in echelon form.

It might be the case that swapping rows, Type II operations, are required to reduce to echelon form, but Type III operations are

never required, they will be required for the further reduction to reduced row echelon form, discussed below. ♢

For an echelon matrix:

- If there are any rows of 0's they must occur at the bottom.
- The first non-zero entry in each non-zero row is called a *pivot*. These are the entries in red in the example above. If only Type I and Type II operations are used, then the pivots can be used to compute the determinant for square matrices.
- The column that a pivot occurs in is called a *pivot column*.
- If the matrix correspond to a system of equations, then the variables corresponding to pivot columns are *pivot variable*, the remaining variable are called *free variables*.
- If the matrix corresponds to an augmented matrix for a system of equations and there is a pivot in the final column, then the system has no solution, since this row would correspond to $0x_1 + 0x_2 + \cdots + 0x_n = p$ where p is the pivot. If this happens, the system is called *inconsistent* otherwise it is *consistent*. For example
$$\left[\begin{array}{cccc:c} 1 & -2 & 1 & 2 & 1 \\ 0 & 2 & -4 & 1 & 3 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$
would be the augmented matrix for an inconsistent system of linear equations.
- If the matrix corresponds to an augmented matrix for a consistent system of equations, then back substitution will solve the system where the pivot variables are given as functions of the free variables. For example
$$\left[\begin{array}{cccc:c} 1 & -2 & 1 & 2 & 1 \\ 0 & 2 & -4 & 1 & 3 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$
is consistent with free variable x_3, all other variables are pivots. By back substitution
$$\begin{aligned} x_4 &= 1 \\ 2x_2 - 4x_3 + x_4 = 3 \Rightarrow 2x_2 - 4x_3 + 1 = 3 \Rightarrow x_2 &= 1 + 2x_3 \\ x_1 - 2x_2 + x_3 + 2x_4 = 1 \Rightarrow x_1 - (2 + 4x_3) + x_3 + 2 = 1 \Rightarrow x_1 &= 1 + 3x_3 \end{aligned}$$
So if x_3 is assigned the value t (x_3 is free so any value may be assigned to x_3), then the solution is $\begin{bmatrix} 1+3t \\ 1+2t \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix}$.

Example 2 Given the following matrix which is in echelon form

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

The pivot columns are in red and the pivots are boxed. If this was the coefficient matrix for a system of equations, then the pivot variables are x_2, x_4, and x_6 and the free variables are x_1, x_3, and x_5. ♢

Two matrices A and B are called *row equivalent* if B arises from A by a series of elementary row operations.

Problem 58 Show that row equivalence is an equivalence relation on M_{mn}, that is

- (Symmetry) If A row equivalent to B, then B is row equivalent to A.
- (Reflexivity) A is row equivalent to A.
- (Transitivity) If A row equivalent to B and B row equivalent to C, then A is row equivalent to C.

For the first item argue that the inverse of a Type I operation is a Type I operation and similarly for Type II and Type III.

✎

See Problem 122 for a nice characterization of row equivalence.

Warning!

A matrix is row equivalent to many matrices in row echelon form.

However the following holds:

If A and B are row equivalent matrices in row echelon form, then A and B have the same pivot columns.

So while there are many row echelon matrices equivalent to a matrix A, all of them have the same pivot columns. Because of this fact we define *the pivot columns of* A to be the pivot columns of B for any echelon matrix which is row equivalent to A.

Define *the rank of* A, denoted $\mathrm{rk}(A)$, to be the number of pivot columns of A.

Define the *nullity of* A, denoted $\mathrm{nullity}(A)$, to be $n - \mathrm{rk}(A)$, where A is $m \times n$. If A is the coefficient matrix for a system of m linear equations in n unknowns, then $\mathrm{nullity}(A)$ is the number of free variables and $\mathrm{rk}(A)$ is the number of pivot variables. Notice that rank, and nullity, are properties which are invariant under elementary row operations.

Problem 59 For each of the following matrices, A, find an REF of A and find the pivot columns of A.

(a)

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

(b)

$$A = \begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 6 & 12 & 0 & 6 & -6 \\ 3 & 6 & 1 & 9 & 0 \\ 2 & 4 & 0 & 2 & -2 \end{bmatrix}$$

The back substitution and solving for the pivots in terms of the free variables can be accomplished via continuing elimination to reach a *reduced echelon form* RREF. A matrix A is in reduced echelon form iff

(1) A is in echelon form.

(2) The only non-zero entry in a pivot column is the pivot.

(3) Each pivot has value 1.

Given an REF matrix A elementary row operations can be used to reach an RREF of A by killing all non-pivot entries in a pivot column using Type II operations and then dividing the pivot rows, a Type III operation, so as to make the pivots have value 1. This is done from right to left whereas the operations to transform A to REF are carried out left to right.

There is only one B in RREF row equivalent to A

Due to this fact write $\mathrm{rref}(A)$ as the unique RREF matrix row equivalent to A.

Problem 60 For each matrix in Problem 59 find the unique matrix $\mathrm{rref}(A)$.

Given a matrix in RREF row equivalent to the augmented matrix of a system of linear equations, it is easy to read off the solution. The RREF matrix A

$$\left[\begin{array}{ccccc:c} 0 & 1 & 2 & 0 & -3 & 4 \\ 0 & 0 & 0 & 1 & 2 & -5 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

corresponds to the system

$$\begin{aligned} x_2 + 2x_3 - 3x_5 = 4 &\Rightarrow x_2 = 4 - 2x_3 + 3x_5 \\ x_4 + 2x_5 = -5 &\Rightarrow x_4 = -5 - 2x_5 \end{aligned}$$

Note that the last row is basically ignored. Assigning values to the free variables we have $x_1 = c_1$, $x_3 = c_2$, and $x_5 = c_3$. The solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} r \\ 4 - 2s + 3t \\ s \\ -5 - 2t \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 0 \\ -5 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} c_1 + \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} c_2 + \begin{bmatrix} 0 \\ 3 \\ 0 \\ -2 \\ 1 \end{bmatrix} c_3$$

The matrix above can be interpreted as

$$\begin{bmatrix} x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -5 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} x_1 - \begin{bmatrix} 2 \\ 0 \end{bmatrix} x_3 - \begin{bmatrix} -3 \\ 2 \end{bmatrix} x_5$$

so filling in the free variables fives

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 4 \\ 0 \\ -5 \\ 0 \end{bmatrix}}_{\text{specific}} + \underbrace{\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} c_1 + \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} c_2 + \begin{bmatrix} 0 \\ 3 \\ 0 \\ -2 \\ 1 \end{bmatrix} c_3 \right)}_{\text{homogeneous part}}$$

Each system $A\boldsymbol{x} = \boldsymbol{b}$ which has a solution, i.e., the system is consistent, has a "general solution" of the form $\boldsymbol{x}_{\text{specific}} + \boldsymbol{x}_{\text{null}}$ where $\boldsymbol{x}_{\text{specific}}$ is **any** solution to $A\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{x}_{\text{null}}$ is a description of all of the solutions to the *homogeneous system* $A\boldsymbol{x} = \boldsymbol{0}$.

Consider the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$. Suppose there are $k = \text{nullity}(A)$ free variables, $x_{i_1}, \ldots, x_{i_k}$. For $1 \le j \le k$, let $\boldsymbol{x}_j^{\text{special}}$ be the solution to $A\boldsymbol{x} = \boldsymbol{0}$ where $x_{i_s} = \begin{cases} 0 & \text{if } s \neq j \\ 1 & \text{otherwise} \end{cases}$, these are the k *special solutions to* $A\boldsymbol{x} = \boldsymbol{0}$.

Given an equation $A\boldsymbol{x} = \boldsymbol{b}$ and any solution $\boldsymbol{x}_{\text{specific}}$ all other solutions are of the form $\boldsymbol{x}_{\text{specific}} + \boldsymbol{x}_{\text{null}}$, where

$$\boldsymbol{x}_{\text{null}} = c_1 \boldsymbol{x}_1^{\text{special}} + \cdots + c_k \boldsymbol{x}_k^{\text{special}}$$

is a linear combination of the k special solutions corresponding to the k free variables.

For the example above we have

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 4 \\ 0 \\ -5 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_{\text{specific}}} + \underbrace{\left(\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_1^{\text{special}}} c_1 + \underbrace{\begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_2^{\text{special}}} c_2 + \underbrace{\begin{bmatrix} 0 \\ 3 \\ 0 \\ -2 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_3^{\text{special}}} c_3 \right)}_{\boldsymbol{x}_{\text{null}}}$$

Problem 61 For each system $A\boldsymbol{x} = \boldsymbol{b}$ find the "general solution" for each of the following systems, note the systems are the same as in Problem 59.

(a)

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 3 & 1 & 5 & -3 & 4 \\ -2 & -3 & -1 & -5 & 6 & -1 \\ 4 & 6 & -1 & 7 & -4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -5 \\ 8 \\ 7 \end{bmatrix}$$

(b)

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 6 & 12 & 0 & 6 & -6 \\ 3 & 6 & 1 & 9 & 0 \\ 2 & 4 & 0 & 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -7 \\ 6 \\ 15 \\ 2 \end{bmatrix}$$

For A an $m \times n$ matrix:

- Say that A has *full column rank* iff $\mathrm{rk}(A) = n$. In this case $A\boldsymbol{x} = \boldsymbol{b}$ has either 1 or 0 solutions.

- Say that A has *full row rank* iff $\mathrm{rk}(A) = m$. In this case there are either one, if $m = n$, or infinitely many, if $n > m$, many solutions.
- If $\mathrm{rk}(A) < \min\{m, n\}$, then $A\boldsymbol{x} = \boldsymbol{b}$ has either 0, 1, or infinitely many solutions.

Problem 62 For $A \in M_{mn}$, view $A\boldsymbol{x} \mapsto \boldsymbol{y}$ as defining the function $L_A : \mathbb{R}^n \to \mathbb{R}^m$. Show

(a) L_A is *one-one*, i.e., $A\boldsymbol{x} = A\boldsymbol{y} \Rightarrow \boldsymbol{x} = \boldsymbol{y}$ iff $\mathrm{rk}(A) = n$ (full column rank).

(b) L_A is *onto*, i.e., for all $\boldsymbol{b} \in \mathbb{R}^m$, there is $\boldsymbol{x} \in \mathbb{R}^n$ such that $A\boldsymbol{x} = \boldsymbol{b}$, iff $\mathrm{rk}(A) = m$ (full row rank).

This leads to the following:

Problem 63 Show that for a square $n \times n$ matrix A, the following are equivalent:

(i) A is invertible.

(ii) L_A is onto.

(iii) A has a right inverse.

(iv) $\mathrm{rk}(A) = n$.

(v) $\mathrm{rref}(A) = I$.

(vi) L_A is one-one.

(vii) A has a left inverse.

This shows that if A has a left inverse, i.e., a matrix B so that $AB = I$, then $B = A^{-1}$. Similarly for right inverse.

Problem 64 For square matrices A and B, show that AB is invertible iff both A and B are invertible.

2.2 Elementary Matrices

Each elementary row operation performed on a $m \times n$ matrix A results from multiplication on the left by a corresponding $m \times m$ *elementary matrix* E. To compute E simply perform the given elementary row operation on I_m.

Problem 65 For A a 3×4 matrix. Write the elementary matrix, E_i, for each of the following elementary row operations op_i. Using the fact that the inverse of an elementary row operation is another elementary row operation of the same type, also write E_i^{-1}, for each operation op_i below.

(op_1) $\mathrm{row}_1 \Leftarrow \mathrm{row}_1 + 3\,\mathrm{row}_2$.

(op_2) $\mathrm{row}_2 \Leftarrow \mathrm{row}_2 - \mathrm{row}_3$.

(op$_3$) $\text{row}_2 \Leftrightarrow \text{row}_1$.

(op$_4$) $\text{row}_3 \Leftarrow \frac{1}{2}\,\text{row}_3$.

Write the matrix that results from applying each elementary operation in order, this is the product of the matrices found above. Write the inverse of this matrix as well.

What would change if A the original matrix is 4×3, or 3×2?

Problem 66 Show that A is invertible iff A is a product of elementary matrices.

Problem 67 Show that A and B in M_{mn} are row equivalent iff $B = PA$ for some invertible matrix $P \in M_{mm}$. This easily gives that row equivalence is an equivalence relation.

2.2.1 Finding A^{-1}

Since A is invertible iff $EA = I$ for some E that is the product of elementary matrices we can find E by starting with $\left[\, A \,\vdots\, I \,\right]$ and performing row operations to transform A into $\text{rref}(A) = I$, giving matrix E. In this way get $E\left[\, A \,\vdots\, I \,\right] = \left[\, EA \,\vdots\, EI \,\right] = \left[\, I \,\vdots\, E \,\right]$.

Problem 68 Find A^{-1} using Gauss-Jordan elimination for

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

Problem 69 Recall a Hermitian matrix (real or complex) is *positive definite* iff $\boldsymbol{x}^* A \boldsymbol{x} > 0$ for all (appropriate) $\boldsymbol{x} \neq \boldsymbol{0}$. Show that a symmetric (Hermitian) positive definite matrix is always invertible by showing that $A\boldsymbol{x} = \boldsymbol{0}$ has the unique solution $\boldsymbol{x} = \boldsymbol{0}$.[9]

Note: $z = \boldsymbol{x}^* A \boldsymbol{x}$ is real since
$z^* = (\boldsymbol{x}^* A \boldsymbol{x})^* = \boldsymbol{x}^* A^* \boldsymbol{x}^{**} = \boldsymbol{x}^* A \boldsymbol{x} = z$.

2.3 LU-decomposition

Returning to the example from page 57 each of the elimination steps used a Type I elementary row operation. For each $i > j$, there was an (i, j) step of the elimination which looked like $\text{row}_i \Leftarrow \text{row}_i - L_{ij}\,\text{row}_j$ for some scalar L_{ij}. These operations are

[9] More generally, for $\boldsymbol{x} \neq \boldsymbol{0}$, $A\boldsymbol{x} = \lambda \boldsymbol{x} \Rightarrow \lambda > 0$.

given by elementary matrices. In the example above we have:

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \quad E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/2 & 1 \end{bmatrix}$$

You can double check that

$$E_{32}E_{31}E_{21}A = U \text{ and } E_{32}E_{31}E_{21}\boldsymbol{b} = \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix}$$

To solve $A\boldsymbol{x} = \boldsymbol{b}$ we are now left solving $U\boldsymbol{x} = \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix}$ which we solve simply with back substitution giving $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Each elementary matrix is clearly invertible since the inverse of

"replace $\text{row}_i(A)$ by $\text{row}_i(A) + \alpha\, \text{row}_j(A)$"

is simply

"replace $\text{row}_i(A)$ by $\text{row}_i(A) - \alpha\, \text{row}_j(A)$"

So

$$E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \quad E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/2 & 1 \end{bmatrix}$$

We also know from Problem 52 that

$$(E_{32}E_{31}E_{21})^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}$$

We can compute the product on the right by simply applying the corresponding elementary row operations to I and the result is the **lower triangular matrix**

$$L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \boxed{1} & 1 & 0 \\ \boxed{2} & \boxed{-1/2} & 1 \end{bmatrix}$$

Note that in the (i, j) elimination step we performed

replace $\text{row}_i(A)$ by $\text{row}_i(A) - L_{ij}\, \text{row}_j(A)$

So L can be trivially recovered from the elimination procedure.

We have the **LU** decomposition of A:

$$E_{32}E_{31}E_{21}A = U \text{ so } A = (E_{32}E_{31}E_{21})^{-1}U = LU$$

Now instead of solving the original $A\boldsymbol{x} = \boldsymbol{b}$ we can solve $L\boldsymbol{c} = \boldsymbol{b}$ and then $U\boldsymbol{x} = \boldsymbol{c}$, since this is equivalent to $A\boldsymbol{x} = L(U\boldsymbol{x}) = L\boldsymbol{c} = \boldsymbol{b}$. So consider solving

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}}_{L\boldsymbol{c}=\boldsymbol{b}} \text{ and } \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}}_{U\boldsymbol{x}=\boldsymbol{c}}$$

For $L\boldsymbol{c} = \boldsymbol{b}$ use forward substitution:

$$\begin{aligned} c_1 &= 3 \\ c_2 &= 1 - c_1 = -2 \\ c_3 &= 1 - 2c_1 + 1/2\, c_3 = 1 - 6 - 1 = -6 \end{aligned}$$

Now we are left solving $U\boldsymbol{x} = \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix}$:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ -6 \end{bmatrix}$$

which we do by back substitution as before to get $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Summary: When solving the system $A\boldsymbol{x} = \boldsymbol{b}$ (for an $n \times n$ coefficient matrix A assuming the system has a unique solution) we can *decompose* A into the product LU where L is lower triangular and U is upper triangular and then solve $A\boldsymbol{x} = \boldsymbol{b}$ by solving for $\boldsymbol{c}$ so that $L\boldsymbol{c} = \boldsymbol{b}$ and then solving $U\boldsymbol{x} = \boldsymbol{c}$, since L and U are diagonal these two steps are easy. Use forward substitution for $L\boldsymbol{c} = \boldsymbol{b}$ and back substitution for $U\boldsymbol{x} = \boldsymbol{c}$.

Problem 70 Find the LU decomposition of A and use it to solve $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 0 & 3 \\ -2 & -7 & 5 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ -16 \end{bmatrix}$$

The same technique may be applied to a square matrix, without a full set of pivots, or more generally, to a rectangular matrix. That is, one can attempt using just Type I operations to get a decomposition $A = LU$ where if A is $m \times n$, then L is lower triangular $m \times m$ with 1's on the diagonal and U is in echelon form. It this is possible, this LU-decomposition of A is "essentially unique", it is actually unique when A is square $n \times n$ and has n-pivots.

Uniqueness. If an invertible matrix A admits an LU with 1's on the diagonal of L, then this decomposition is unique.

Proof. Suppose $A = L_1U_1 = L_2U_2$. Since A is invertible so are all the other matrices, so you get $L_2^{-1}L_1 = U_2U_1^{-1}$. The inverse of an upper/lower triangular matrix is also upper/lower triangular so the left hand side is lower triangular and the right hand side is upper triangular, thus both sides are in fact diagonal. The matrices L_1 and L_2^{-1} have 1's on the diagonal and hence $L_2^{-1}L_1$ has 1's on the diagonal (see Problem 46). Thus $L_2^{-1}L_1 = I$ and hence $L_1 = L_2$, similarly $U_1 = U_2$. □

Existence. For a square $n \times n$ matrix A let A_k be the submatrix formed by taking the upper left $k \times k$ corner. A has an LU decomposition iff each A_k $(k = 1, ..., n)$ is invertible.

Fact: If A is Hermitian and positive definite, then A has an LU decomposition. (The point is that a Hermitian positive definite matrix is invertible (see Problem 69) and each of the A_k's will also be such a matrix. (See Problem 45.))

2.4 $A = LDU$ decomposition

If A admits an LU decomposition with full pivots, then U can be factored into $U = DU'$ where D is the diagonal matrix of pivots and U' has 1's in the diagonal.

Problem 71 For the matrix

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

from Problem 70 find the LDU factorization of A.

✎

Problem 72 Suppose A is invertible, symmetric, and has an LDU decomposition. Show that $U = L^T$, so that $A = LDL^T$. (Similarly, if A is Hermitian, then $A = LDL^*$, recall that the diagonal elements of a Hermitian matrix are real and so the entries in D, i.e., the pivots, are real.)

Problem 73 Find the decomposition $A = LDL^T$ for

$$A = \begin{bmatrix} -1 & -2 & 1 \\ -2 & -2 & 6 \\ 1 & 6 & 8 \end{bmatrix}$$

Problem 74 If A is Hermitian, positive definite, then A has an LU decomposition, and so $A = LDL^*$. Show that $A = KK^*$ for some matrix (unique) K.

2.5 $PA = LU$ decomposition

As explained above row exchanges might be required to convert A into an upper triangular matrix U. A row exchange is a type of

permutation of rows which are given by permutation matrices. For example we want to permute rows as follows:

$$\begin{bmatrix} \text{row}_1(A) \\ \hdashline \text{row}_2(A) \\ \hdashline \text{row}_3(A) \end{bmatrix} \Rightarrow \begin{bmatrix} \text{row}_3(A) \\ \hdashline \text{row}_1(A) \\ \hdashline \text{row}_2(A) \end{bmatrix}$$

Then the permutation matrix is

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

You can check that this is not a row exchange, but can be accomplished by two row exchanges. Permutation matrices enjoy many nice properties:

- If P and P' are permutation matrices, then so is PP'.
- $P^{-1} = P^T$.
- The rows/columns of P are mutually orthogonal.

Problem 75 One notation for a permutation of $\{1, 2, 3\}$ would be $(1, 2)$ meaning $1 \to 2 \to 1$, or $(1, 2, 3)$ meaning $1 \to 2 \to 3 \to 1$. These can be composed, for example, $(1, 2)(2, 3)$ would be $3 \to 1$, $2 \to 3$, and $1 \to 2$, finally, $()$ is just the identity. There are $3! = 6$ permutations of $\{1, 2, 3\}$:

$$\begin{aligned} \sigma_0 &= () & \sigma_1 &= (1, 2) \\ \sigma_2 &= (2, 3) & \sigma_3 &= (1, 3) \\ \sigma_4 &= (3, 2, 1) & \sigma_5 &= (1, 2, 3) \end{aligned}$$

Let P_i be the matrix corresponding to σ_i, so $P_0 = I$. Make a multiplication table for the P_i's and verify that $P_i^{-1} = P_i^T$.

Problem 76 Show that for a permutation matrix P, $P^{-1} = P^T$ as follows:

(a) Show $(P\boldsymbol{x}) \cdot (P\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$. (Matrices with this property are called *unitary*.)

(b) Rewriting you get $\boldsymbol{y}^T P^T P\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{x}$. Since this is true for **all** $\boldsymbol{y}$ **show that** $P^T P\boldsymbol{x} = \boldsymbol{x}$ for **all** $\boldsymbol{x}$.

(c) Argue that since $(P^T P)\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x}$ it follows that $P^T P = I$!

There are many ways to find P and then the resulting $PA = LU$. Some methods (partial pivoting) minimize round off error; the method I suggest works, but is not optimal in any particular way. Simply put we perform elimination is a column, then look for an appropriate pivot and then swap rows if necessary in A.

Suppose you start with an $n \times n$ matrix, and assume that A is invertible so that the process does not get stuck.

Step 0. Make sure there is a non-zero entry in the $(1, 1)$ position. This gets us started with a $A_0 = P_0 A$.

Step 1. Perform elimination on the first column. This gives the first column of L_1. Find the least i such that the $(i, 2)$ entry is non-zero, this will become the next pivot. A_0 has been reduced to A_1' which has 0's in the first column below the first pivot. Now apply $P_{(i,2)}$ to A_0 and eliminate as before, this will result in $A_1 = P_{(i,2)}A_1'$. Also make the appropriate change in L_1 to give L_1. Let $P_1 = P_{(i,2)}P_0$. Elimination in the first column of $P_1 A$ results in A_1 and A_1 has a non-zero $(2, 2)$ entry.

Step 2. Now start eliminating below the $(2, 2)$ element of A_1. ETC.

This continues to stage $n - 1$ and results in $P = P_{n-1}$ and $L = L_{n-1}$ so that $PA = LU$.

The above procedure assumes that a pivot could be found at each stage, i.e., that PA has full pivots, if this fails we need to modify the procedure a bit and will in section 3.

Problem 77 Find a $PA = LU$ decomposition and solve $A\boldsymbol{x} = \boldsymbol{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 1 & 2 & 1 \\ 2 & 3 & 4 & 5 \\ 2 & 2 & 2 & 1 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 6 \\ 2 \\ 9 \\ 2 \end{bmatrix}$$

Note you want to solve $PA\boldsymbol{x} = P\boldsymbol{b} = \boldsymbol{d}$, now solve $LU\boldsymbol{x} = \boldsymbol{d}$ in the usual fashion.

Problem 78 Find a $PA = LU$ decomposition for

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

2.5.1 Extending decomposition to rectangular matrices

The $PA = LU$ decomposition extends to producing REF and additional decomposition produces RREF.

Steps to produce rref(A):

$$A \underset{\text{Type II}}{\Rightarrow} PA \underset{\text{Type I}}{\Rightarrow} \underbrace{U}_{\substack{\text{echelon}\\ PA=LU}} \underset{\text{Type I}}{\Rightarrow} D \underset{\text{Type III}}{\Rightarrow} R = \text{rref}(A)$$

Example As an example consider finding rref(A) for

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 4 & -1 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix}$$

Begin with $P_{(1,2)}$ giving

$$\begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix}$$

Now eliminate in the first column

$$\text{row}_2(A) \Leftarrow \text{row}_2(A) - 0\,\text{row}_1(A) \tag{2,1}$$
$$\text{row}_3(A) \Leftarrow \text{row}_3(A) - 2\,\text{row}_1(A) \tag{3,1}$$
$$\text{row}_4(A) \Leftarrow \text{row}_4(A) - 3\,\text{row}_1(A) \tag{4,1}$$

This results in

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & ? & 1 & 0 \\ 3 & ? & ? & 1 \end{bmatrix} \quad U_1 = \begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -2 & 4 \end{bmatrix}$$

Now apply $P_{(3,4)}$ to get the next pivot in the correct position, the result is in echelon form so adjusting L accordingly we have

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Notice the 1,3,4,5 columns are pivot columns and hence the pivot variables are x_1, x_3, x_4, x_5. The free variable is x_2.

Here we have

$$P = P_{(3,4)} P_{(1,2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and $PA = LU$:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 4 & -1 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Now eliminate from right to left upward above pivots on U

$$\text{row}_3 \Leftarrow \text{row}_3 - 2\,\text{row}_4 \qquad (3,5)$$
$$\text{row}_2 \Leftarrow \text{row}_2 - {}^3\!/_2\,\text{row}_4 \qquad (2,5)$$
$$\text{row}_1 \Leftarrow \text{row}_1 - (-{}^1\!/_2)\,\text{row}_4 \qquad (1,5)$$

This results in

$$\hat{U} = \begin{bmatrix} 1 & ? & ? & -{}^1\!/_2 \\ 0 & 1 & ? & {}^3\!/_2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 1 & 2 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Note the column in $\hat{U}$ is formed by simply dividing the corresponding column of U by the appropriate pivot and this continues to hold.

Eliminating above the $(3,4)$ entry is done with

$$\text{row}_2 \Leftarrow \text{row}_2 - (-1)\,\text{row}_3 \qquad (2,4)$$
$$\text{row}_1 \Leftarrow \text{row}_1 - (-2)\,\text{row}_3 \qquad (1,4)$$

This results in

$$\hat{U} = \begin{bmatrix} 1 & ? & -2 & -1/2 \\ 0 & 1 & -1 & 3/2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

The final step is

$$\text{row}_1 \Leftarrow \text{row}_1 - 2\,\text{row}_2 \tag{1,3}$$

This results in

$$\hat{U} = \begin{bmatrix} 1 & 2 & -2 & -1/2 \\ 0 & 1 & -1 & 3/2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

So now we have $U = \hat{U}D$ so $PA = L\hat{U}D$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 4 & -1 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix} =$$

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

The final step is to transform the pivots of D into 1's so we have

$$\text{row}_3 \Leftarrow 1/{-2}\,\text{row}_3$$
$$\text{row}_4 \Leftarrow 1/2\,\text{row}_4$$

giving

$$\hat{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \text{ and } R = \text{rref}(A) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $PA = L\hat{U}\hat{D}R$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 4 & -1 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix} =$$

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

So we have

$$A \underset{P}{\Rightarrow} PA \underset{L^{-1}}{\Rightarrow} \underbrace{U}_{\substack{\text{echelon}\\ PA=LU}} \underset{\hat{U}^{-1}}{\Rightarrow} D \underset{\hat{D}^{-1}}{\Rightarrow} R = \text{rref}(A)$$

This expresses $R = \hat{D}^{-1}\hat{U}^{-1}L^{-1}PA$ where L^{-1} is just the product of type I elementary matrices used to reduce PA to U, and $\hat{U}^{-1}$ is the product of type I matrices used to reduce U to D and finally $\hat{D}^{-1}$ is the product of type III matrices used to reduce D to R.

What is described here is more than we are usually interested in, the following captures the most relevant aspects:

$$A \xrightarrow{\text{row}_1 \Leftrightarrow \text{row}_2} \begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 2 & 2 & 4 & 8 & 0 \\ 3 & 3 & 6 & 10 & 1 \end{bmatrix} \xrightarrow[\text{row}_4 \Leftarrow \text{row}_4 - 3\,\text{row}_1]{\text{row}_3 \Leftarrow \text{row}_3 - 2\,\text{row}_1} \begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -2 & 4 \end{bmatrix}$$

$$\xrightarrow{\text{row}_3 \Leftrightarrow \text{row}_4} \underbrace{\begin{bmatrix} 1 & 1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}}_{U} \xrightarrow[\substack{\text{row}_2 \Leftarrow \text{row}_2 - 3/2\,\text{row}_4 \\ \text{row}_1 \Leftarrow \text{row}_1 + 1/2\,\text{row}_4}]{\text{row}_3 \Leftarrow \text{row}_3 - 2\,\text{row}_4} \begin{bmatrix} 1 & 1 & 2 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$\xrightarrow[\text{row}_1 \Leftarrow \text{row}_1 + 2\,\text{row}_3]{\text{row}_2 \Leftarrow \text{row}_2 + \text{row}_3} \begin{bmatrix} 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \xrightarrow{\text{row}_1 \Leftarrow \text{row}_1 - 2\,\text{row}_2} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$\xrightarrow[\text{row}_4 \Leftarrow 1/2\,\text{row}_4]{\text{row}_3 \Leftarrow -1/2\,\text{row}_3} \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{R}$$

$\diamondsuit$

Problem 79 Produce rref(A) for each of the following:

(a)

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

(b)

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

3 Vector spaces and subspaces

The algebraic properties satisfied by $\mathbb{R}^n$ under the operations of vector addition and scalar multiplication are as follows:

Algebraic properties of vector spaces

Axiom 1: (Associativity of vector addition) For all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$

$$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) =$$

Axiom 2: (Existence of additive identity) There is some $\boldsymbol{e}$ such that for all $\boldsymbol{u}$

$$\boldsymbol{u} + \boldsymbol{e} = \boldsymbol{e} + \boldsymbol{u} =$$

Axiom 3: (Existence of additive inverses) For all $\boldsymbol{u}$ there is a $\boldsymbol{v}$ such that

$$\boldsymbol{u} + \boldsymbol{v} =$$

Axiom 4: (Commutativity of vector addition) For all vectors $\boldsymbol{u}$ and $\boldsymbol{v}$

$$\boldsymbol{u} + \boldsymbol{v} =$$

Axiom 5: (Associativity of scalar multiplication) For all scalars α, β and all vectors $\boldsymbol{v}$

$$\alpha(\beta \boldsymbol{v}) =$$

Axiom 6: (Distributivity) For all scalars α, β and all vectors $\boldsymbol{u}, \boldsymbol{v}$

$$(\alpha + \beta)\boldsymbol{v} =$$

$$\alpha(\boldsymbol{u} + \boldsymbol{v}) =$$

Axiom 7: (Unitary property) For all vectors $\boldsymbol{v}$

$$1\boldsymbol{v} =$$

Any set of objects satisfying the above eight properties is called a *vector space over* $\mathbb{R}$. The corresponding notion of a vector space over the scalar field $\mathbb{C}$ is defined in a completely analogous manner.

Remark For those going on to *Abstract Algebra*, the first 3 axioms state that $(\mathbb{R}^n, +, \mathbf{0})$ is a *group*, the 4^{th} axiom states that this is an *abelian* or *commutative* group. The remaining axioms state that the field $(\mathbb{R}, +, \times, 0, 1)$ *acts* on $(\mathbb{R}^n, +, \mathbf{0})$ in a structure preserving way. Notice we are using the same symbol "+" for addition in the scalar field and for vector addition, in this class we will continue to be sloppy in this way. ♢

We are most interested in the vector spaces $\mathbb{R}^n$ (or $\mathbb{C}^n$) and associated subspaces, however, in this section, for completeness, we will consider a sampling of more general vector spaces.

Some additional properties can be derived from the above

- The additive identity is unique we call it $\mathbf{0}$.
- $(0)\boldsymbol{v} = \mathbf{0}$ for all $\boldsymbol{v}$.
- The additive inverse of $\boldsymbol{v}$ is unique and we call it $-\boldsymbol{v}$.
- $-\boldsymbol{v} = (-1)\boldsymbol{v}$.

Problem 80 (a) Using just the axioms show that the additive identity and the additive inverses are unique, and hence we write $-\boldsymbol{v}$ for the additive inverse of $\boldsymbol{v}$. We also denote the additive identity as $\mathbf{0}$.

(b) Using just the axioms show that $(0)\boldsymbol{v} = \mathbf{0}$.

(c) Using just the axioms show that $(-1)\boldsymbol{v} = -\boldsymbol{v}$.

Problem 81 Show that M_{22} is a vector space. (The same obviously hold for M_{mn}.)

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

Problem 82 Show that the set of infinite sequences of reals, $\mathbb{R}^{\mathbb{N}}$, $\boldsymbol{x} = \langle x_i \rangle_{i=0}^{\infty}$ forms a vector space. (Here, $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of *natural numbers*.) Vector addition is defined by $\langle x_i \rangle_{i=0}^{\infty} + \langle y_i \rangle_{i=0}^{\infty} = \langle x_i + y_i \rangle_{i=0}^{\infty}$ and scalar multiplication by $\alpha \langle x_i \rangle_{i=0}^{\infty} = \langle \alpha x_i \rangle_{i=0}^{\infty}$.

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

Problem 83 Generalizing from the preceding, show that the set of all real valued functions, $\mathbb{R}^X$, $f : X \to \mathbb{R}$ (for X some fixed set), forms a vector space. Here we define vector addition by $(f+g)(x) = f(x) + g(x)$ and scalar multiplication by $(\alpha f)(x) = \alpha f(x)$.

Verify that $\mathbb{R}^n$, $\mathbb{R}^{\mathbb{N}}$, M_{mn} are all of the form $\mathbb{R}^X$ for appropriately chosen X. Many vector spaces are of this form.

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

Problem 84 Show that the set of all n^{th} degree polynomials with real coefficients in the variable x, $P_n[x]$ is a vector space. Here vector addition is defined by
$\sum_{i=0}^{n} \alpha_i x^i + \sum_{i=0}^{n} \beta_i x^i = \sum_{i=0}^{n} (\alpha_i + \beta_i) x^i$ and scalar multiplication by $\alpha \sum_{i=0}^{n} \beta_i x^i = \sum_{i=0}^{n} (\alpha\beta_i) x^i$.

Similarly, the set of all polynomials with real coefficients, $P[x]$ is a vector space as is the set of *formal power series*, i.e.,
$p(x) = \sum_{i=0}^{\infty} a_i x^i$.

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

Problem 85 Show the following are vector spaces:

(a) $\mathbb{R}^+ = (0, \infty)$ with *vector addition* given by $r \oplus s = rs$ (standard multiplication in $\mathbb{R}$) and scalar multiplication $\alpha \odot r = r^\alpha$.[10] (The additional symbols are necessary as we must distinguish multiplication of two scalars from scalar multiplication of a scalar and a vector, for example $2 \odot 3 = 3^2$ is clear, whereas, if we just wrote (2)(3) it is not clear if this is multiplication of two scalars or if it is the scalar product of a scalar and a vector?)

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

[10] The map $x \mapsto a^x$ is linear! This is because $a^{\alpha x+\beta y} = (a^x)^\alpha (a^y)^\beta = (\alpha \odot a^x) \oplus (\beta \odot a^y)$.

(b) $\mathbb{R}^2$ with vector addition defined by $(x_1, x_2) \oplus (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1)$ and scalar multiplication defined by $r \odot (x_1, x_2) = (rx_1 + r - 1, rx_2 + r - 1)$.

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

3.1 Subspaces

Generally we are not interested in particular vectors, but rather entire *spaces* of vectors. Recall that a *vector space* is any set satisfying the vector space properties on page 87.

Mostly we are interested in subspaces of $\mathbb{R}^n$ (or $\mathbb{C}^n$), however, in this section, for completeness, we will consider a small collection of more general subspaces.

For V a vector space, $W \subseteq V$ is a *subspace* of V iff W is itself a vector space.

Problem 86 (**Important!**) $\emptyset \neq W \subseteq \mathbb{R}^n$ is a subspace iff W is closed under linear combinations.

To check that W is a subspace it suffices to check the following two conditions for $\boldsymbol{u}, \boldsymbol{v} \in W$ and $\alpha \in \mathbb{R}$

(1) $\boldsymbol{u} + \boldsymbol{v} \in W$.

(2) $\alpha \boldsymbol{u} \in W$.

Prove this!!!

Axiom 1.

Axiom 2.

Axiom 3.

Axiom 4.

Axiom 5.

Axiom 6.

Axiom 7.

There are always two *trivial subspaces* of a vector space V, namely, $W = V$ and $W = \{\mathbf{0}\}$.

Problem 87 Describe, geometrically, all possible subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$.

Problem 88 (1) Show by example, say in $\mathbb{R}^2$ or $\mathbb{R}^3$, that if W_1 and W_2 are subspaces of V, then $W_1 \cup W_2$ need **not** be a subspace of V.

(2) Show that if W_1 and W_2 are subspaces of V, then $W_1 \cap W_2$ is a subspace of V.

(3) Show that if W_1 and W_2 is a subspace of V, then $W_1 + W_2 = \{\boldsymbol{w}_1 + \boldsymbol{w}_2 \,|\, \boldsymbol{w}_1 \in W_1 \,\&\, \boldsymbol{w}_2 \in W_2\}$ is a subspace of V.

(4) (**Important!!**) Show that if $S \subseteq V$, then the set of all linear combination of elements from S,

$$\operatorname{Span}(S) \stackrel{\text{df}}{=} \left\{ \sum_{i=1}^{n} \alpha_i \boldsymbol{w}_i \,\middle|\, \alpha_i \in \mathbb{R} \,\&\, \boldsymbol{w}_i \in S \right\}$$

is the smallest subspace of V containing S.

Problem 89 For W, W_1, W_2 subspaces of a vector space V, write $W = W_1 \oplus W_2$ to mean $W = W_1 + W_2$ and $W_2 \cap W_2 = \{\mathbf{0}\}$ and say W *is the direct sum of* W_1 *and* W_2. Suppose $W = W_1 \oplus W_2$ show that

(a) For $\boldsymbol{w} \in W$, there is exactly one $\boldsymbol{w}_1$ and exactly one $\boldsymbol{w}_2$ such that $\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2$, in fact, $\boldsymbol{w}_2 = \boldsymbol{w} - \boldsymbol{w}_1$.

(b) Define the *projection of* $\boldsymbol{w}$ *into* W_1 *along* W_2 by $\text{proj}_{W_2,W_1}(\boldsymbol{w})$ to be the unique $\boldsymbol{w}_1 \in W_1$ so that there is some $\boldsymbol{w}_2 \in W_2$ with $\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2$. Similarly define *the projection into* W_2 *along* W_1, $\text{proj}_{W_1,W_2} : W \to W_2$.[11] Show that the two functions proj_{W_2,W_1} and proj_{W_1,W_2} are linear.

Problem 90 Show that if $W \subseteq \mathbb{R}^n$ and $W \neq \mathbb{R}^n$, then $W^\perp = \{\boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{u} \perp W\}$ is a subspace of $\mathbb{R}^n$. Here $\boldsymbol{u} \perp W$ means that $\boldsymbol{u} \perp \boldsymbol{w}$ for all $\boldsymbol{w} \in W$. Later we will show $\mathbb{R}^n = W \oplus W^\perp$.

[11] Notice that $\text{proj}_{W_2,W_1}(\boldsymbol{w})$ is defined with respect to W_1 and W_2 in the sense that if $W = W_1 \oplus W_2 = W_1 \oplus W_3$, then $\text{proj}_{W_2,W_1} \neq \text{proj}_{W_3,W_1}$. To see this consider W_1, W_2, and W_3 to be three distinct lines through the origin in $\mathbb{R}^2$, so that $\mathbb{R}^2 = W_1 \oplus W_2 = W_1 \oplus W_3$. This also indicates that W_2 is not determined by W_1, that is, there is no sense in which $W_2 = W - W_1$.

Problem 91 In $\mathbb{R}^2$

(a) Why is the *quarter plane* consisting of all $\boldsymbol{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ with $a, b > 0$ not a subspace?

(b) Why is the *two quarter plane* consisting of all $\boldsymbol{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ where a and b have the same sign not a subspace?

Problem 92 Geometrically and algebraically describe

$\operatorname{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$ in $\mathbb{R}^3$.

Problem 93 Show that the set of diagonal $m \times n$ matrices is a subspace of M_{mn}.

Problem 94 Show that the set of upper triangular $n \times n$ matrices is a subspace of M_{nn}. (Similarly for lower triangular.)

Problem 95 Show that the symmetric and skew-symmetric $n \times n$ matrices form subspaces of M_{nn}.

Problem 96 Show that the set of all *absolutely convergent series*, i.e., the set of all $\langle x_i \rangle_{i=0}^{\infty}$ such that $\sum_{i=0}^{\infty} |x_i| < \infty$ is a subspace of $\mathbb{R}^{\mathbb{N}}$. This subspace is called ℓ^1.[12]

[12] Other interesting spaces are th ℓ^p spaces, these satisfy $\sum_{i=0}^{\infty} |x_i|^p, \infty$ for $1 \leq p < \infty$.

Problem 97 Show that the set $C(\mathbb{R}, \mathbb{R})$ of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$.[13]

3.2 Basis and dimension

A set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is *linearly dependent* iff there is some *non-trivial*[14]linear combination of the vectors in $\mathcal{B}$ that gives the $\mathbf{0}$ vector, otherwise $\mathcal{B}$ is *linearly independent*.

[13]Similarly, the set $C^n(\mathbb{R}, \mathbb{R})$ of all functions with continuous n^{th} derivatives.

Problem 98 Show that $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is linearly independent iff for all i,
$\boldsymbol{v}_i \notin \mathrm{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_k\}) = \mathrm{Span}(\mathcal{B} \setminus \{\boldsymbol{v}_i\})$.

Problem 99 Show that $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is linearly independent iff for each $\boldsymbol{v} \in \mathrm{Span}(\mathcal{B})$, there is unique $\boldsymbol{a} = (a_1, \ldots, a_k) \in \mathbb{R}^k$ such that $\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_k\boldsymbol{v}_k$, denote this as $\boldsymbol{a}^{\mathcal{B}}$. This unique tuple is denotes $[\boldsymbol{v}]_{\mathcal{B}}$ and is called the *component representation of* $\boldsymbol{v}$ *with respect to* $\mathcal{B}$. Show that the map $L : \mathrm{Span}(\mathcal{B}) \to \mathbb{R}^k$ given by $L : \boldsymbol{v} \mapsto [\boldsymbol{v}]_{\mathcal{B}}$ is linear, one-one, and onto. Show that
$L^{-1} : \mathbb{R}^n \to \mathrm{Span}(\mathcal{B})$ is given by $L^{-1} : \boldsymbol{a} \mapsto \boldsymbol{a}^{\mathcal{B}}$, so $([\boldsymbol{v}]_{\mathcal{B}})^{\mathcal{B}} = \boldsymbol{v}$ and $[\boldsymbol{a}^{\mathcal{B}}]_{\mathcal{B}} = \boldsymbol{a}$.

A set of vectors S *spans* a subspace W of a vector space V iff $\mathrm{Span}(S) = W$. Given a set S that spans W it is possible to throw away some vectors to get $S_0 \subseteq S$ such that $\mathrm{Span}(S_0) = \mathrm{Span}(S)$ and S_0 is linearly independent. Conversely if S is a linearly independent set of vectors in W then it is possible to find $S_1 \supseteq S$ so that S_1 is linearly independent and spans W. The next few problems support these claims.

Problem 100 Show that $\mathrm{Span}(S_0) \subseteq \mathrm{Span}(S_1)$ iff for each $\boldsymbol{v} \in S_0$, $\boldsymbol{v} \in \mathrm{Span}(S_1)$.

Problem 101 Suppose S is a *maximal linearly independent* set of

[14] A linear combination $\sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$ is trivial iff $\alpha_i = 0$ for all i, that is $\sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i = \sum_{i=1}^{n} (0) \boldsymbol{v}_i = \mathbf{0}$.

vectors in W, that is, S is linearly independent and for all $\boldsymbol{v} \in W \setminus S$, $S \cup \{\boldsymbol{v}\}$ is linearly dependent. Show that $\operatorname{Span}(S) = W$.

Problem 102 Show that if S is a *minimal spanning set* for W, that is, $\operatorname{Span}(S) = W$ and $\operatorname{Span}(S \setminus \{\boldsymbol{v}\}) \neq W$ for all $\boldsymbol{v} \in S$, then S is linearly independent.

Problem 103 Suppose S spans W and $\boldsymbol{v} \in W$, show that there is $\boldsymbol{u} \in S$ so that $S' = (S \setminus \{\boldsymbol{u}\}) \cup \{\boldsymbol{v}\}$ spans W.

Problem 104 Suppose $S_0 = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$ is a linearly independent set of vectors in W and $S_1 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l\}$ spans W. Show that $k \leq l$.

Hint: Essentially use the preceding problem to form $S_1^i = \left(S_1^{i-1} \setminus \{\boldsymbol{v}\}\right) \cup \{\boldsymbol{w}_i\}$. You must show that $\boldsymbol{v}$ can be chosen not to be one of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{i-1}$. At each stage you maintain $\operatorname{Span}(S_1^i) = W$.

Problem 105 Use the preceding problem to argue that if W has a finite spanning set, then any two linearly independent spanning sets must be of the same size.

Thus a spanning set for W has *enough* vectors to generate W, and a linearly independent set doesn't have *too many*, i.e., there is no redundant information. A set that is linearly independent and spanning is *just right*, not too many vectors nor too few.

A *basis* for W is a linearly independent spanning set. Three important facts are:

- Every linearly independent set can be filled out (by adding vectors) to form a basis.
- Every spanning set can be cut down (by deleting vectors) to form a basis.
- Every basis for W has the same size, called the *dimension* of W, and denoted $\dim(W)$.

We have proved this in the exercises above for W that can be spanned by a finite number of vectors, i.e., *finite dimensional vector spaces.* The infinite dimensional case requires a bit more machinery.

Problem 106 Find a basis for each of the following spaces:

(a) M_{33} - the space of all 3×3 matrices.

(b) The set of symmetric 3×3 matrices. Temporarily call this subspace S_{33}.

(c) The space of all lower triangular 3×3 matrices. Temporarily call this subspace L_{33}.

(d) The space $S_{33} \cap L_{33}$.

(e) The space $S_{33} + L_{33}$.

Problem 107 Show that for U, W subspaces of a finite dimensional vector space V

$$\boxed{\dim(U) + \dim(W) = \dim(U + W) + \dim(U \cap W)}$$

A special case of the preceding is when $U \cap W = \{\mathbf{0}\}$. In this case we know

- $\dim(U+W) = \dim(U) + \dim(W)$
- If $\mathcal{B}_U$ is a basis for U and $\mathcal{B}_W$ is a basis for W, then $\mathcal{B}_U \cup \mathcal{B}_W$ is a basis for $U+W$.

Recall $V = U \oplus W$ to denotes $V = U + W$ together with the fact that $U \cap W = \{\mathbf{0}\}$, in this case V is the *direct sum of U and W* and U, W form a *decomposition of V*.

A vector space V can be decomposed into more than two subspaces. V is the *direct sum* of subspaces $V_1, \ldots, V_k$, written $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, is defined to mean:

- $V = V_1 + V_2 + \cdots + V_k$, i.e., every vector $\boldsymbol{v}$ can be written as $\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k$ where $\boldsymbol{v}_i \in V_i$.
- $V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_k) = \{\mathbf{0}\}$, i.e., the decomposition $\boldsymbol{v} = \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k$ with $\boldsymbol{v}_i \in V_i$ is unique.

Problem 108 Let $E_i = \operatorname{Span}\{\boldsymbol{e}_i^n\}$ be the i^{th} axis in $\mathbb{R}^n$. Show that $\mathbb{R}^n = E_1 \oplus E_2 \oplus \cdots \oplus E_n$.

Problem 109 Show that $\mathbb{R}^2 = U \oplus W$ where $U = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$ and $W = \operatorname{Span}\left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$. Generalize to explain why $\mathbb{R}^2 = U + W$ where U and W are any two non-colinear lines through the origin.

The next problem generalizes this to $\mathbb{R}^3$.

Problem 110 Show that $\mathbb{R}^3 = U_1 \oplus U_2 \oplus U_3$ where

$U_1 = \text{Span}\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}\right\}$, $U_2 = \text{Span}\left\{\begin{bmatrix}-1\\1\\0\end{bmatrix}\right\}$, and $U_3 = \text{Span}\left\{\begin{bmatrix}0\\1\\-1\end{bmatrix}\right\}$.
Generalize to explain why $\mathbb{R}^3 = U_1 \oplus U_2 \oplus U_3$ where U_1, U_2 and U_3 are any three non-coplanar lines through the origin.

Finally generalize to an arbitrary vector space.

Problem 111 Show that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are linearly independent iff $\text{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}) = \text{Span}(\{\boldsymbol{v}_1\}) \oplus \cdots \oplus \text{Span}(\{\boldsymbol{v}_k\})$.

The next problem just has you think about some important infinite dimensional subspaces of some of the vector spaces we have considered previously.

Problem 112 Show that the following are infinite dimensional: (There is a natural progression in these five examples.)

(a) $P[x]$ - the set of all polynomials in x with real coefficients. Find a basis for this $P[x]$.

(b) $\mathbb{R}^{<\infty} = \{(x_i)_{i=0}^{\infty} \,|\, x_i \in \mathbb{R} \;\&\;$ for some N, $x_i = 0$ for all $i > N\}$. This is essentially $\bigcup_{n=1}^{\infty} \mathbb{R}^n$. Clearly $\mathbb{R}^{<\infty}$ and $P[x]$ are "essentially the same".

(c) ℓ^1 - the set of all $\boldsymbol{\alpha} = \langle \alpha_i \rangle_{i=0}^{\infty}$ such that $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. (This is the vector space of all absolutely convergent sequences.)[15]

(d) $\mathcal{P}$ - the set of all power series $p(x) = \sum_{i=0}^{\infty} c_i x^i$ whose radius of convergence is $R_p \geq 1$. Recall $|x| < R_p \Rightarrow \langle c_i x^i \rangle \in \ell^1$.

(e) $C^{\infty}((-1,1), \mathbb{R})$ - the set of functions $f : (-1,1) \to \mathbb{R}$ with derivatives of all orders.

[15] You might have fun showing that there is no *countably infinite basis!*. It is clear that if one allowed *infinite linear combinations*, then setting $\boldsymbol{e}_i^{\infty}$ to be the member of ℓ^1 with 0's everywhere except the i^{th} entry, we have $\langle \alpha_i \rangle = \sum_{i=0}^{\infty} \alpha_i \boldsymbol{e}_i^{\infty}$. These sums however intrinsically involve limits so we are moving away from pure algebra and into analysis. What is true is that there is a countable subset D of ℓ^1 such that any element of ℓ^1 is approximated arbitrarily good by some element in D.

3.3 Change of bases and similar matrices

Given a set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ for $\mathbb{R}^n$ and $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$ in $\mathbb{R}^k$ we defined

$$\boldsymbol{x}^{\mathcal{B}} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_k\boldsymbol{v}_k$$

Conversely, if $\mathcal{B}$ is linearly independent and $\boldsymbol{v} \in \operatorname{Span}(\mathcal{B})$, then

$$\boldsymbol{v}_{\mathcal{B}} = \boldsymbol{x} \in \mathbb{R}^k \leftrightarrow \boldsymbol{x}^{\mathcal{B}} = \boldsymbol{v}.$$

$\boldsymbol{v}_{\mathcal{B}}$ is the $\mathcal{B}$-coordinate representation of $\boldsymbol{v}$.

Notice that for $\mathcal{B}$ a basis for $\mathbb{R}^n$, $(\boldsymbol{x}^{\mathcal{B}})_{\mathcal{B}} = \boldsymbol{x}$ and $(\boldsymbol{x}_{\mathcal{B}})^{\mathcal{B}} = \boldsymbol{x}$.

Clearly if we set B to be the matrix whose columns are the vectors in $\mathcal{B}$, that is, $B = \left[\boldsymbol{v}_1 \,\vdots\, \boldsymbol{v}_2 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right]$, then

$$\boldsymbol{x}^{\mathcal{B}} = B\boldsymbol{x}.$$

So $\boldsymbol{x} \to B\boldsymbol{x}$ converts $\mathcal{B}$-coordinates into standard coordinates.

If $\mathcal{B}$ is a basis for $\mathbb{R}^n$, then

$$\boldsymbol{a}^{\mathcal{B}} = B\boldsymbol{a} = \boldsymbol{x} \leftrightarrow \boldsymbol{a} = \boldsymbol{x}_{\mathcal{B}} = B^{-1}\boldsymbol{x}.$$

Thus B^{-1} converts standard coordinates into $\mathcal{B}$-coordinates.

Thus B and B^{-1} are change of basis matrices and the corresponding transformations $\boldsymbol{x} \to \boldsymbol{x}^{\mathcal{B}} = B\boldsymbol{x}$ and $\boldsymbol{x} \to \boldsymbol{x}_{\mathcal{B}} = B^{-1}\boldsymbol{x}$ are change of bases transformations.

Example 3 Consider the following two bases for $\mathbb{R}^3$:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\} \text{ and } \mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

We have

$$\begin{bmatrix}1\\1\\1\end{bmatrix}^{\mathcal{B}} = \begin{bmatrix}1&3&1\\2&2&3\\3&1&2\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}5\\7\\6\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\end{bmatrix}^{\mathcal{C}} = \begin{bmatrix}1&0&1\\0&1&1\\1&1&0\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\2\\2\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\end{bmatrix}_{\mathcal{B}} = \left(\begin{bmatrix}1&3&1\\2&2&3\\3&1&2\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1/12&-5/12&7/12\\5/12&-1/12&-1/12\\-1/3&2/3&-1/3\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1/4\\1/4\\0\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\end{bmatrix}_{\mathcal{C}} = \left(\begin{bmatrix}1&0&1\\0&1&1\\1&1&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1/2&-1/2&1/2\\-1/2&1/2&1/2\\1/2&1/2&-1/2\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1/2\\1/2\\1/2\end{bmatrix}$$

So the $\mathcal{B}$-coordinate representation of $\begin{bmatrix}1\\1\\1\end{bmatrix}$ is $\begin{bmatrix}1/4\\1/4\\0\end{bmatrix}$.

To go from a $\mathcal{B}$ coordinate representation, $\boldsymbol{x}$, to a $\mathcal{C}$ coordinate representation, $\boldsymbol{y}$, first go from $\mathcal{B}$ to standard and then from standard to $\mathcal{C}$, this is given by

$$\boldsymbol{y} = C^{-1}B\boldsymbol{x}$$

For example

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

Here $\begin{bmatrix}1/4\\1/4\\0\end{bmatrix}$ is the $\mathcal{B}$-coordinate representation of $\begin{bmatrix}1\\1\\1\end{bmatrix}$ and $\begin{bmatrix}1/2\\1/2\\1/2\end{bmatrix}$ is the $\mathcal{C}$ representation of $\begin{bmatrix}1\\1\\1\end{bmatrix}$. ♢

Problem 113 Let $\mathcal{B} = \left\{\begin{bmatrix}1\\2\\-2\end{bmatrix}, \begin{bmatrix}3\\0\\1\end{bmatrix}\right\}$. Let $W = \operatorname{Span}(\mathcal{B})$.

(a) Find a basis, $\mathcal{C}$, for W consisting of orthonormal vectors.

(b) Find $\begin{bmatrix} 2 \\ -1 \end{bmatrix}^{\mathcal{B}}$.

(c) Find $\begin{bmatrix} 4 \\ -10 \\ 13 \end{bmatrix}_{\mathcal{B}}$.

(d) Find the matrix that converts elements of W into their $\mathcal{B}$ coordinate representation.

(e) Find the matrix that converts $\mathcal{B}$ coordinate representations for elements of W into $\mathcal{C}$ coordinate representations.

(f) Notice that $\left\{ \begin{bmatrix} u \\ v \end{bmatrix}^{\mathcal{C}} \middle| \, u^2 + v^2 = 1 \right\}$ gives the unit circle in the plane W at the origin. Convert this into an equation for the same unit circle in terms of the variables x, y, z. Notice $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = C \begin{bmatrix} u \\ v \end{bmatrix}$ and $\begin{bmatrix} u \\ v \end{bmatrix} = P_{\mathcal{C}} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ for $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in W$. Use this together with $u^2 + v^2 = 1$ to get a system of equations describing the circle:

In the preceding we used that if $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is linearly independent, then

$$P_{\mathcal{B}} = (B^T B)^{-1} B^T$$

gives the conversion from standard basis representation for

elements of Span($\mathcal{B}$) to $\mathcal{B}$-coordinates. It is important to know $B^T B$ is invertible for this.

Notice that if $\mathcal{B}$ is a basis for V, then the above reduces to

$$P_{\mathcal{B}} = B^{-1}$$

since $(B^T B)^{-1} B^T = B^{-1}(B^T)^{-1} B^T = B^{-1}$.

Problem 114 Show that if $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is linearly independent, and $B = \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_k\right]$, then $B^* B$ is invertible.

Notice that this shows that in order to solve $A\boldsymbol{x} = \boldsymbol{b}$ where A has linearly independent columns and $\boldsymbol{b}$ is in the range of $\boldsymbol{x} \mapsto A\boldsymbol{x}$, we may simply compute

$$\boldsymbol{x} = (A^T A)^{-1} A^T \boldsymbol{b}.$$

3.4 Fundamental subspaces and general solution to $A\mathbf{x} = \mathbf{b}$

There are four fundamental subspaces associated to an $m \times n$ matrix. Here we want to study those subspaces by applying Gauss elimination to find bases for these subspaces, then further understand the general solution to $A\boldsymbol{x} = \boldsymbol{b}$ for arbitrary A.

3.4.1 Column space and Null space

The *column space* of a matrix $A \in M_{mn}$ is the smallest subspace of $\mathbb{R}^m$ containing all of the columns of A, formally

$$\mathrm{CS}(A) \stackrel{\mathrm{df}}{=} \mathrm{Span}\left(\mathrm{col}_1(A), \dots, \mathrm{col}_n(A)\right) \subseteq \mathbb{R}^m$$

Problem 115 Describe $\mathrm{CS}(A)$ for $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$. Is this all of $\mathbb{R}^4$? Are all of the columns necessary? What is a "geometric" description of the subspace?

Problem 116 With A as above, write down some $\boldsymbol{b}$'s so that $A\boldsymbol{x} = \boldsymbol{b}$ has a solution. How could you describe the set of all such $\boldsymbol{b}$'s?

The point of the previous two problems was to get you to see that

$$A\boldsymbol{x} = \boldsymbol{b} \text{ has a solution } \Leftrightarrow \boldsymbol{b} \in \mathrm{CS}(A).$$

The goal is not only simply to solve $A\boldsymbol{x} = \boldsymbol{b}$ for a specific $\boldsymbol{b}$, but more generally to describe the set (space) of vectors $\boldsymbol{b}$ such that $A\boldsymbol{x} = \boldsymbol{b}$ has a solution. So we would like to be able to describe $\mathrm{CS}(A)$ in a "nice" way. Part of what Problem 115 indicates is that $\mathrm{CS}(A)$ only depends on the first two of the columns (or more generally any two).

Given a system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$, the associated *homogeneous system* of equations is $A\boldsymbol{x} = \boldsymbol{b}$. The homogeneous system **always** has a solution. (**What is it?**) The set of solutions to the homogeneous system is our next subspace.

The second fundamental subspace associated to an $m \times n$ matrix A is the *null space of* A, this is the set of all solutions of $A\boldsymbol{x} = \mathbf{0}$, formally,

$$\mathrm{NS}(A) \stackrel{\text{df}}{=} \{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \mathbf{0}\}$$

Problem 117 Show that $\mathrm{NS}(A)$ is a subspace of $\mathbb{R}^n$ for $A \in M_{mn}$.

Problem 118 For A as in Problem 115, that is $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$, characterize geometrically and algebraically, the null space of A.

Problem 119 Let A be as in the preceding problem. Find, by inspection, a specific solution to,

$$A\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \boldsymbol{b}.$$

Call the solution you found $\boldsymbol{c}$. Show that every element of $\boldsymbol{c} + \mathrm{NS}(A)$ is also a solution to $A\boldsymbol{x} = \boldsymbol{b}$. Argue that this is all of the solutions. Describe, geometrically, the set of solutions to $A\boldsymbol{x} = \boldsymbol{b}$. (This is what we will call the *general solution*.)

For vectors in $\mathbb{R}^m$, we can phrase linear independence in terms of matrices by noting that the following are equivalent for $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ in $\mathbb{R}^m$:

- $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are linearly independent.
- $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution.
- $\mathrm{NS}(A) = \{\mathbf{0}\}$, that is, $\mathrm{NS}(A)$ is trivial.

where here A is the $m \times k$ matrix with $\mathrm{row}_i(A) = \boldsymbol{v}_i$, that is,

$$A = \left[\, \boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_k \,\right].$$

Problem 120 Show that the following are equivalent for a square $n \times n$ matrix A

(i) A is invertible.

Continued from Problem 63.

(viii) $\mathrm{CS}(A) = \mathbb{R}^n$.

(ix) $\mathrm{NS}(A) = \{\mathbf{0}\}$.

(x) The columns of A form a basis for $\mathbb{R}^n$.

(xi) The rows of A form a basis for $\mathbb{R}^n$.

The *row space of* A, denoted $\mathrm{RS}(A)$, is the span of the rows of A, that is, for $A \in M_{mn}$

$$\mathrm{RS}(A) \overset{\mathrm{df}}{=} \mathrm{Span}\left(\mathrm{row}_1(A), \ldots, \mathrm{row}_m(A)\right).$$

Clearly, $\mathrm{RS}(A) = \mathrm{CS}(A^T)$.

Problem 121 Show that $\mathrm{NS}(A) = \mathrm{RS}(A)^{\perp}$.

The last of the four fundamental spaces associated to A is $\mathrm{NS}(A^T)$, this is called the *left nullspace* since

$$\boldsymbol{x} \in \mathrm{NS}(A^T) \Leftrightarrow A^T\boldsymbol{x} = \mathbf{0} \Leftrightarrow \boldsymbol{x}^T A = \mathbf{0}^T \text{(left nullspace)}.$$

Recall matrices A and B *row equivalent* if B results from performing elementary row operations to A.

Problem 122 Show that the following are equivalent for matrices of the same size:

(1) A and B are row equivalent.

(2) $\mathrm{RS}(A) = \mathrm{RS}(B)$.

(3) $\mathrm{NS}(A) = \mathrm{NS}(B)$.

Problem 123 Prove that if B and C are in echelon form and are row equivalent to A, then B and C have the same pivot columns. Hint: Consider the first column which is a pivot column of one and not the other and derive a contradiction.

If A is in row echelon form, then:

(i) The (non-zero) rows are linearly independent and form a basis for the row space.

(ii) The pivot columns form a basis for the column space.

From this it follows that $\dim(\mathrm{RS}(A)) = \#$ of pivots and this is called the *rank of* A, denoted $\mathrm{rank}(A)$. If A, B, and C are row equivalent with B and C being row echelon, then since $\mathrm{RS}(A) = \mathrm{RS}(B) = \mathrm{RS}(A)$ we know that B and C have the same number of pivot columns.

$$\boxed{\mathrm{rank}(A) = \# \text{ of pivots of } A}$$

Problem 124 Prove the uniqueness of $\mathrm{rref}(A)$.

Hint: Suppose B and C are RREF matrices row equivalent to A. We know that B and C have exactly the same pivots from Problem 123.

If A is $m \times n$ so that $A\boldsymbol{x} = \boldsymbol{b}$ has n variables, then

$$\boxed{\#(\text{pivot variables}) + \#(\text{free variables}) = n.}$$

3.4.2 The four fundamental subspaces

Recall that $\mathrm{RS}(A) = \mathrm{RS}(B)$ if A and B are row equivalent, so

- If A is row equivalent to B and B is in echelon form, then the non-zero rows of B are a basis for $\mathrm{RS}(A)$.
- $\dim(\mathrm{RS}(A)) = \mathrm{rank}(A)$.

In particular the "canonical basis" for $\mathrm{RS}(A)$ consists of the non-zero rows of $\mathrm{rref}(A)$.

What happens with the columns is even more interesting. First of all notice that generally $\mathrm{CS}(A) \neq \mathrm{CS}(B)$ for row equivalent A and B, so the answer is different than for row space. The important fact is:

If $A, B \in M_{mn}$ are row equivalent and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, then

$$\begin{aligned}\mathrm{col}_{i_1}(A), \ldots, \mathrm{col}_{i_k}(A) \text{ are linearly independent } \Leftrightarrow \\ \mathrm{col}_{i_1}(B), \ldots, \mathrm{col}_{i_k}(B) \text{ are linearly independent}\end{aligned}$$

Proof. Let $A = \left[\boldsymbol{v}_1 \,\vdots\, \boldsymbol{v}_2 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right]$ and set $\hat{A} = \left[\boldsymbol{v}_{i_1} \,\vdots\, \boldsymbol{v}_{i_2} \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_{i_k}\right]$. Since A and B are row equivalent, there is invertible $m \times m$ matrix P so that $PA = B$, but
$PA = P\left[\boldsymbol{v}_1 \,\vdots\, \boldsymbol{v}_2 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right] = \left[P\boldsymbol{v}_1 \,\vdots\, P\boldsymbol{v}_2 \,\vdots\, \cdots \,\vdots\, P\boldsymbol{v}_n\right] = B$, so $P\boldsymbol{v}_i$ is the i^{th} column of B.

Let $\hat{B} = P\hat{A}$, then clearly,

$$\hat{B}\boldsymbol{x} = \mathbf{0} \Leftrightarrow P\hat{A}\boldsymbol{x} = \mathbf{0} \Leftrightarrow \hat{A}\boldsymbol{x} = P^{-1}\mathbf{0} = \mathbf{0}$$

So $\mathrm{NS}(\hat{A}) = \mathrm{NS}(\hat{B})$. We have

$$\begin{aligned}\text{The columns of } \hat{A} \text{ are linearly independent } &\Leftrightarrow \\ \mathrm{NS}(\hat{A}) = \{\mathbf{0}\} &\Leftrightarrow \\ \mathrm{NS}(\hat{B}) = \{\mathbf{0}\} &\Leftrightarrow \\ \text{The columns of } \hat{B} \text{ are linearly independent}\end{aligned}$$

□

In particular

$$\boxed{\text{The pivot columns of } A \text{ form a basis for } \mathrm{CS}(A).}$$

As a consequence of this

$$\boxed{\dim(\mathrm{CS}(A)) = \dim(\mathrm{RS}(A)) = \mathrm{rank}(A)}$$

Next consider $\mathrm{NS}(A)$ for $A \in M_{mn}$ of rank r. As with $\mathrm{RS}(A)$ it is the case that $\mathrm{NS}(A) = \mathrm{NS}(B)$ whenever A and B are row equivalent, thus $\mathrm{NS}(A) = \mathrm{NS}(\mathrm{rref}(A))$. By permuting the columns of $\mathrm{rref}(A)$ (equivalently permuting the variables) we can view $R = \mathrm{rref}(A)$ as having the form

$$\left[\begin{array}{c:c} I_r & F \\ \hdashline O & O \end{array}\right]$$

and $R\boldsymbol{x} = \mathbf{0}$ becomes

$$\left[\begin{array}{c:c} I_r & F \\ \hdashline O & O \end{array}\right] \left[\begin{array}{c} \boldsymbol{x}_{\text{pivot}} \\ \hdashline \boldsymbol{x}_{\text{free}} \end{array}\right] = \left[\begin{array}{c} \mathbf{0}_r \\ \hdashline \mathbf{0}_{n-r} \end{array}\right]$$

that is

$$\boldsymbol{x}_{\text{pivot}} = -F\boldsymbol{x}_{\text{free}}$$

This can be expressed as

$$\left[\begin{array}{c} \boldsymbol{x}_{\text{pivot}} \\ \hdashline \boldsymbol{x}_{\text{free}} \end{array}\right] = \left[\begin{array}{c} -F \\ \hdashline I_{n-r} \end{array}\right] \left[\boldsymbol{x}_{\text{free}}\right]$$

The matrix that corresponds to permuting the rows of $\left[\begin{array}{c} -F \\ \hdashline I_{n-r} \end{array}\right]$ back to the standard order of the variables is the *null space matrix* and its columns clearly span null space.

Another way to view the basis for $\mathrm{NS}(A)$ is as follows: Consider the $n - r$ assignments to the free variables where in each case one variable is assigned the value 1 and the rest are assigned the value 0 (similar to the standard basis for $\mathbb{R}^{n-r}$). All possible assignments to the free variables arise as linear combinations of

these *special assignments*. To each special assignment there is an associated *special solutions*, namely, those of the form

$$\boldsymbol{x}_{\text{free}} = \boldsymbol{e}_i^{n-r} \text{ and } \boldsymbol{x}_{\text{pivot}} = -\text{col}_i(F)$$

These $n-r$ special solutions are linearly independent elements of $\text{NS}(A)$. So $\dim(\text{NS}(A)) \geq n-r$, that the special solutions span $\text{NS}(A)$ follows from the fact that $\text{NS}(A) + \text{RS}(A) = \mathbb{R}^n$, which is shown in Problem 125 below. This shows $\dim(\text{NS}(A)) = n-r$. Before arguing this, first let's look an example.

Example Say $R = \text{rref}(A)$ is

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

So the pivot variables are x_1, x_3, x_6 and the rest are free, thus

$$\boldsymbol{x}_{\text{pivot}} = \begin{bmatrix} x_1 \\ x_3 \\ x_6 \end{bmatrix} \text{ and } \boldsymbol{x}_{\text{free}} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_7 \end{bmatrix} \text{ and we view } R\boldsymbol{x} = \mathbf{0} \text{ as}$$

$$\left[\begin{array}{ccc:cccc} 1 & 0 & 0 & 2 & 1 & -1 & 1 \\ 0 & 1 & 0 & 0 & -3 & 2 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_3 \\ x_6 \\ \hdashline x_2 \\ x_4 \\ x_5 \\ x_7 \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hdashline 0 \\ 0 \\ 0 \\ 0 \end{array}\right]$$

so

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

Method 1: (Compute the null space matrix first and then the special solutions.) The null space matrix is the appropriate permutation of rows of

$$\left[\begin{array}{c} -F \\ \hdashline I_{n-r} \end{array}\right] = \left[\begin{array}{cccc} -2 & -1 & 1 & -1 \\ 0 & 3 & -2 & 1 \\ 0 & 0 & 0 & -2 \\ \hdashline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

So the null space matrix is

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

The columns of this matrix are the canonical basis for $\text{NS}(A)$, these are also the special solutions.

Method 2: (Compute the special solutions first and then the null space matrix.) The special solutions are determined by the assignments:

$$\boldsymbol{x}_{\text{free}} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_7 \end{bmatrix} = \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{x}_{\text{pivot}} = \begin{bmatrix} x_1 \\ x_3 \\ x_6 \end{bmatrix} = -\text{col}_1(F) = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{x}_{\text{free}} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_7 \end{bmatrix} = \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{x}_{\text{pivot}} = \begin{bmatrix} x_1 \\ x_3 \\ x_6 \end{bmatrix} = -\text{col}_2(F) = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}$$

$$\boldsymbol{x}_{\text{free}} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_7 \end{bmatrix} = \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{x}_{\text{pivot}} = \begin{bmatrix} x_1 \\ x_3 \\ x_6 \end{bmatrix} = -\text{col}_3(F) = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

$$\boldsymbol{x}_{\text{free}} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_7 \end{bmatrix} = \boldsymbol{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{x}_{\text{pivot}} = \begin{bmatrix} x_1 \\ x_3 \\ x_6 \end{bmatrix} = -\text{col}_4(F) = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}$$

Thus the special solutions and a basis for $\text{NS}(A)$ are

$$\boldsymbol{x}_1^{\text{special}} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}_2^{\text{special}} = \begin{bmatrix} -1 \\ 0 \\ 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}_3^{\text{special}} = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}_4^{\text{special}} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

with a little practice you can read these off without actually permuting the variables, writing F, etc. The *null space matrix* is

the matrix with the $n - r$ special solutions as columns:

$$\left[\boldsymbol{x}_1^{\text{special}} \,\vdots\, \boldsymbol{x}_2^{\text{special}} \,\vdots\, \boldsymbol{x}_3^{\text{special}} \,\vdots\, \boldsymbol{x}_4^{\text{special}}\right] = \begin{bmatrix} -2 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 3 & -2 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

◊

Remark. The i^{th} special solution is the *unique solution* to $A\boldsymbol{x} = \mathbf{0}$ that you get by setting $\boldsymbol{x}_{\text{free}}$ to be $\boldsymbol{e}_i^n$. In other words, it does not make a difference whether $\operatorname{rref}(A)$, A, or any other matrix B that is row equivalent to A, is used to find the special solutions.

Problem 125 Show that $\mathbb{R}^n = \operatorname{RS}(A) \oplus \operatorname{NS}(A)$, that is

- $\operatorname{NS}(A) \cap \operatorname{RS}(A) = \{\mathbf{0}\}$ and
- $\mathbb{R}^n = \operatorname{NS}(A) + \operatorname{RS}(A)$.

✎

The *nullity* of A is the dimension of the null space of A and we have that for A an $m \times n$ matrix

$$\boxed{\operatorname{nullity}(A) + \operatorname{rank}(A) = n}$$

alternatively

$$\boxed{\dim(\operatorname{NS}(A)) + \dim(\operatorname{RS}(A)) = \dim(\operatorname{NS}(A)) + \dim(\operatorname{CS}(A)) = n}.$$

Problem 126 For the matrices in Example on page 82 and Problem 79 find the rank, nullity, and canonical bases for each of $\operatorname{RS}(A)$, $\operatorname{CS}(A)$, and $\operatorname{NS}(A)$.

The final issue is to find a canonical basis for the fourth fundamental subspace, the left nullspace, $\mathrm{NS}(A^T)$.

Method 1: Find $\mathrm{rref}(A^T)$ and then compute the basis for $\mathrm{NS}(A^T)$ as the special solutions to $A^T\boldsymbol{x} = \mathbf{0}$ just as you did for $\mathrm{NS}(A)$.

Method 2: For this we must keep track of the matrix leading to $\mathrm{rref}(A)$, this can be done in a manner similar to the way we found A^{-1} in the invertible case. Set up

$$\left[A_{m\times n} \,\vdots\, I_{m\times m}\right] \xrightarrow{\text{row ops}} \left[R \,\vdots\, E\right],$$

where $R = \mathrm{rref}(A)$ and $EA = R$, that is E is the product of the elementary matrices associated to the row operations used in reducing A to reduced row echelon form.

We are interested in when $\boldsymbol{x}^T A = \mathbf{0}^T$, this shows up here if R has final rows of 0's, suppose $\mathrm{rank}(A) = r < m$, then

$$EA = \begin{bmatrix} \boldsymbol{u}_1^T \\ \hline \vdots \\ \hline \boldsymbol{u}_r^T \\ \hline \boldsymbol{u}_{r+1}^T \\ \hline \vdots \\ \hline \boldsymbol{u}_m^T \end{bmatrix} A = \begin{bmatrix} \boldsymbol{v}_1^T \\ \hline \vdots \\ \hline \boldsymbol{v}_r^T \\ \hline \mathbf{0}^T \\ \hline \vdots \\ \hline \mathbf{0}^T \end{bmatrix} = R$$

The $m - r$ vectors $\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$ are a basis for $\mathrm{NS}(A^T)$.

Fact: These two methods give exactly the same basis for $\mathrm{NS}(A^T)$. The argument for this follows after the next example.

Example 4 For

$$A = \begin{bmatrix} 1/2 & 1 & -1/2 \\ 3 & 6 & 0 \\ 3/2 & 3 & 1/2 \\ 1 & 2 & 0 \end{bmatrix}$$

Find the bases for each of the four fundamental subspaces found by Gauss elimination.

$$\left[A \,\vdots\, I_{4\times 4}\right] = \left[\begin{array}{ccc:cccc} 1/2 & 1 & -\frac{1}{2} & 1 & 0 & 0 & 0 \\ 3 & 6 & 0 & 0 & 1 & 0 & 0 \\ 3/2 & 3 & 1/2 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[\substack{\text{row}_2 \Leftarrow \text{row}_2 - 6\,\text{row}_1 \\ \text{row}_3 \Leftarrow \text{row}_3 - 3\,\text{row}_1 \\ \text{row}_4 \Leftarrow \text{row}_4 - 2\,\text{row}_1}]{} \left[\begin{array}{ccc:cccc} 1/2 & 1 & -1/2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & -6 & 1 & 0 & 0 \\ 0 & 0 & 2 & -3 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[\substack{\text{row}_3 \Leftarrow \text{row}_3 - 2/3\,\text{row}_2 \\ \text{row}_4 \Leftarrow \text{row}_4 - 1/3\,\text{row}_2}]{} \left[\begin{array}{ccc:cccc} 1/2 & 1 & -1/2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & -6 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2/3 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[\text{row}_1 \Leftarrow \text{row}_1 + 1/6\,\text{row}_2]{} \left[\begin{array}{ccc:cccc} 1/2 & 1 & 0 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 3 & -6 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2/3 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[\substack{\text{row}_1 \Leftarrow 2*\text{row}_1 \\ \text{row}_2 \Leftarrow 1/3\,\text{row}_2}]{} \left[\begin{array}{ccc:cccc} 1 & 2 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2/3 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & 0 & 1 \end{array}\right] = \left[R \,\vdots\, E\right]$$

CS(A): The pivot columns are columns 1 and 3 so dim(CS(A)) = 2, moreover, a basis for CS(A) is

$$\begin{bmatrix} 1/2 \\ 3 \\ 3/2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$$

NS(A): dim(NS(A)) = 3 − dim(CS(A)) = 1 and the basis of special solutions to $A\boldsymbol{x} = \mathbf{0}$ is

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

CS(A^T) = RS(A): dim(RS(A)) = dim(CS(A)) = 2 and a basis is

$$\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

NS(A^T): dim(NS(A^T)) = 4 − dim(CS(A^T)) = 2

Method 1: Here $A^T = \begin{bmatrix} 1/2 & 3 & 3/2 & 1 \\ 1 & 6 & 3 & 2 \\ -1/2 & 0 & 1/2 & 0 \end{bmatrix}$ and $\text{rref}(A^T) = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2/3 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. So the special solutions to $A^T\boldsymbol{x} = \mathbf{0}$ are

$$\begin{bmatrix} 1 \\ -2/3 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ -1/3 \\ 0 \\ 1 \end{bmatrix}$$

Method 2: The final two rows of E are exactly the two special solutions just found above.

$\diamondsuit$

The reason that the two methods for finding a basis for $\text{NS}(A^T)$ is as follows: First notice that the first round of elimination, killing non-zeros below pivots, results in $E_1PA = U_1$ where U_1 is in echelon form. Suppose P is unnecessary, or replace A with PA. The matrix E_1 is an $m \times m$ lower triangular matrix of the form

$$\left[\begin{array}{c:c} L & O_{(m-r)\times r} \\ \hdashline B & I_{(m-r)} \end{array}\right]$$

where L is an $r \times r$ lower triangular matrix with 1's on the diagonal and B is an $(m-r) \times r$ matrix such that

$$\left[\begin{array}{c:c} L & O_{(m-r)\times r} \\ \hdashline B & I_{(m-r)} \end{array}\right] \left[\begin{array}{c} A_1 \\ \hdashline A_2 \end{array}\right] = \left[\begin{array}{c} LA_1 \\ \hdashline BA_1 + IA_2 \end{array}\right] = \left[\begin{array}{c} LA_1 \\ \hdashline O_{(m-r)\times n} \end{array}\right]$$

We get here that

$$BA_1 + I_{(m-r)}A_2 = O_{(m-r)\times n}$$

taking transposes gives

$$A_1^T B^T + A_2^T I_{(m-r)} = O_{n\times(m-r)}$$

which can be rewritten as

$$\left[\begin{array}{c} A_1^T \\ \hdashline A_2^T \end{array}\right] \left[\begin{array}{c} B^T \\ \hdashline I_{(m-r)} \end{array}\right] = O_{m\times(m-r)}$$

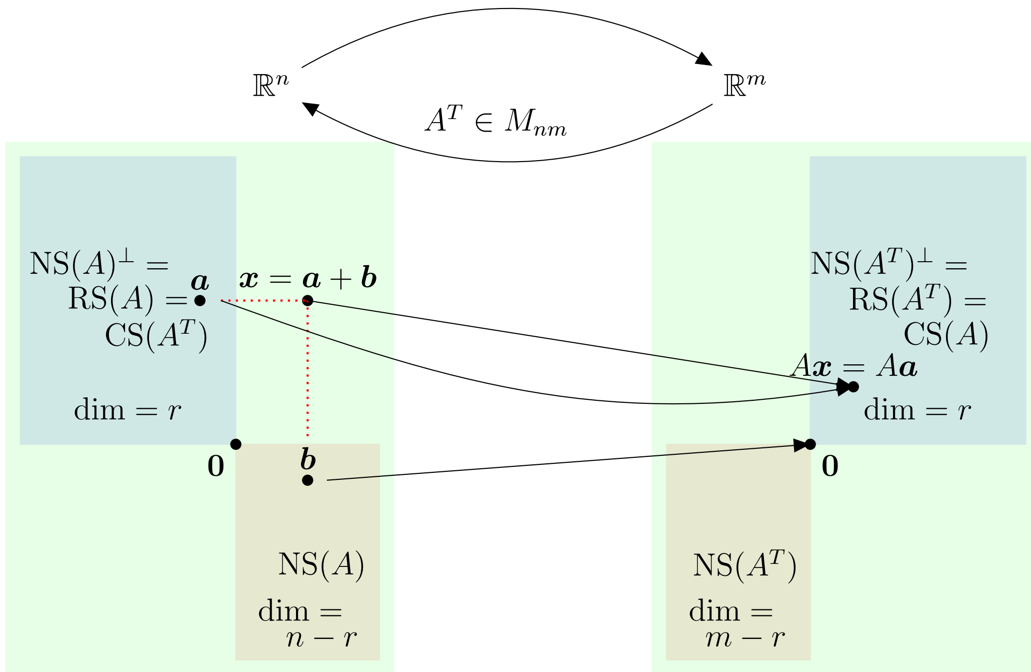
So the matrix $\left[\begin{array}{c} B^T \\ \hdashline I_{(m-r)} \end{array}\right]$ is the matrix of special solutions for $A^T\boldsymbol{x} = \mathbf{0}$.

Problem 127 Find the dimension and canonical basis for each of the four fundamental spaces associated with

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

The four fundamental subspaces of A

$A \in M_{mn} \quad \text{rank}(A) = r$

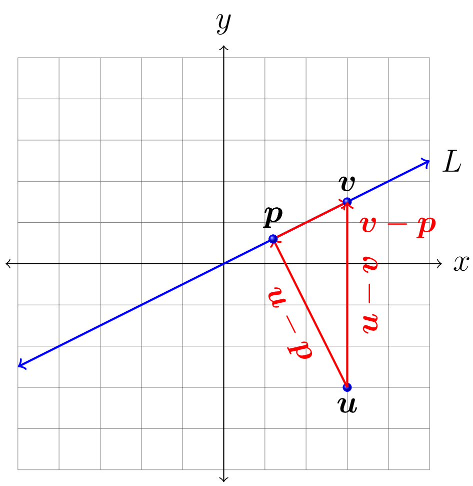


4 Projection operations and least squares solution to $A\mathbf{x} = \mathbf{b}$

Begin with the simplest case of projection onto a one dimensional subspace, that is a line. Let L be a line through the origin in $\mathbb{R}^n$, i.e., a one dimensional subspace of $\mathbb{R}^n$, the projection of $\boldsymbol{u}$ into L is defined as the point $\boldsymbol{p}$ on L so that $\text{dist}(\boldsymbol{u}, \boldsymbol{p})$ is minimal. We will show that such a point exists presently. Let $\{\boldsymbol{a}\}$ be a basis for L, that is $L = \text{Span}\{\boldsymbol{a}\}$. The point/vector $\boldsymbol{p}$ that we seek is $\boldsymbol{p} = x\boldsymbol{a}$ for some scalar x.

Claim: For $\boldsymbol{u}$ a fixed element of $\mathbb{R}^n$ and $\boldsymbol{v}$ in L, $\text{dist}(\boldsymbol{u}, \boldsymbol{p})$ is minimal when $\boldsymbol{u} - \boldsymbol{p}$ is orthogonal to L.

This is intuitively clear as illustrated in the following picture



To see this suppose $\boldsymbol{p} \in L$ and $(\boldsymbol{u} - \boldsymbol{p}) \perp L$, let $\boldsymbol{v} \in L$, then

$$\text{dist}(\boldsymbol{u}, \boldsymbol{v})^2 = \|\boldsymbol{u}-\boldsymbol{v}\|^2 = \|(\boldsymbol{u}-\boldsymbol{p})+(\boldsymbol{p}-\boldsymbol{v})\|^2 = \|\boldsymbol{u}-\boldsymbol{p}\|^2+\|\boldsymbol{p}-\boldsymbol{v}\|^2 \geq \|\boldsymbol{u}-\boldsymbol{p}\|^2 = \text{dist}(\boldsymbol{u}, \boldsymbol{p})^2$$

since as $(\boldsymbol{u} - \boldsymbol{p}) \perp (\boldsymbol{p} - \boldsymbol{v})$ we have the Pythagorean theorem (see Problem 17.)

So we seek $\boldsymbol{p} = x\boldsymbol{a}$ so that $\boldsymbol{a} \cdot (\boldsymbol{u} - x\boldsymbol{a}) = 0$, but then we can solve for x,

$$x = \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\boldsymbol{a} \cdot \boldsymbol{a}} = (\boldsymbol{a}^T\boldsymbol{a})^{-1}\boldsymbol{a}^T\boldsymbol{u}$$

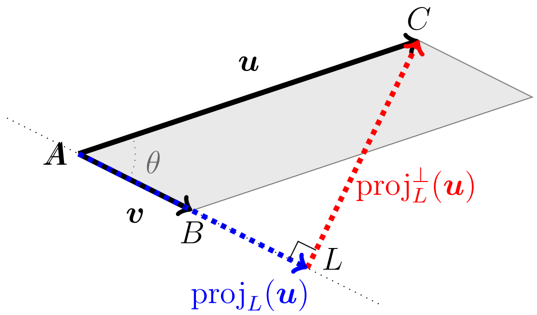
and so we derive the following computation for $\boldsymbol{p}$

$$\boxed{\boldsymbol{p} = \left(\frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\boldsymbol{a} \cdot \boldsymbol{a}}\right)\boldsymbol{a} = [\boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{a})^{-1}\boldsymbol{a}^T]\boldsymbol{u}}$$

This $\boldsymbol{p}$ is the *orthogonal projection of* $\boldsymbol{u}$ *into* L and is denoted $\text{proj}_L(\boldsymbol{u})$. Two things to take note of are:

- It does not matter what $\boldsymbol{a} \in L$ is used and so we can write $\text{proj}_{\boldsymbol{a}}(\boldsymbol{u})$ in place of $\text{proj}_L(\boldsymbol{u})$.

- The orthogonal projection function is linear in $\boldsymbol{u}$, that is, $\operatorname{proj}_L(\alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2) = \alpha_1 \operatorname{proj}_L(\boldsymbol{u}_1) + \alpha_2 \operatorname{proj}_L(\boldsymbol{u}_2)$.



$\operatorname{proj}_L(\boldsymbol{u})$ is linear so there is a matrix P_L satisfying $\operatorname{proj}_L(\boldsymbol{u}) = P_L\boldsymbol{u}$, we can use the formula for projection above to explicitly write P_L as

$$\boxed{P_L = \boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{a})^{-1}\boldsymbol{a}^T}$$

is a rank one matrix which does not depend on choice of $\boldsymbol{a} \in L$.

Notice the range of P_L, that is $\operatorname{CS}(P_L)$, is the 1 dimensional L, while the kernel of P_L, that is $\operatorname{NS}(P_L)$ is $n-1$ dimensional.

The part of $\boldsymbol{u}$ orthogonal to $\boldsymbol{p} = \operatorname{proj}_L(\boldsymbol{u})$ is $\operatorname{proj}_L^{\perp}(\boldsymbol{u}) = \boldsymbol{u} - \operatorname{proj}_L(\boldsymbol{u})$. The matrix for this operation is the rank $(n-1)$ matrix

$$\boxed{P_L^{\perp} \stackrel{\text{df}}{=} P_{L^{\perp}} = I_{n\times n} - P_L.}$$

A discussed in the next section, $P_L^{\perp} = I - P_L$ is also a projection matrix, now projecting into the $(n-1)$-dimensional subspace $\operatorname{NS}(P_L)$ with the property that

$$\operatorname{dist}(\boldsymbol{u}, P_L^{\perp}\boldsymbol{u}) = \min\{\operatorname{dist}(\boldsymbol{u},\boldsymbol{v}) \mid \boldsymbol{v} \in \operatorname{NS}(P_L)\}.$$

This can be illustrated in $\mathbb{R}^3$ where $L^{\perp}$ is a plane, in fact if $L = \operatorname{Span}\left\{\begin{bmatrix}2\\-3\\1\end{bmatrix}\right\}$, then $L^{\perp}$ is the plane $2x - 3y + z = 0$, that is the set of all $\begin{bmatrix}x\\y\\z\end{bmatrix}$ such that $\begin{bmatrix}x\\y\\z\end{bmatrix} \cdot \begin{bmatrix}2\\-3\\1\end{bmatrix} = 0$, that is, $\begin{bmatrix}2 & -3 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = 0$. So $L^{\perp}$ is $\operatorname{NS}\left(\begin{bmatrix}2 & -3 & 1\end{bmatrix}\right)$

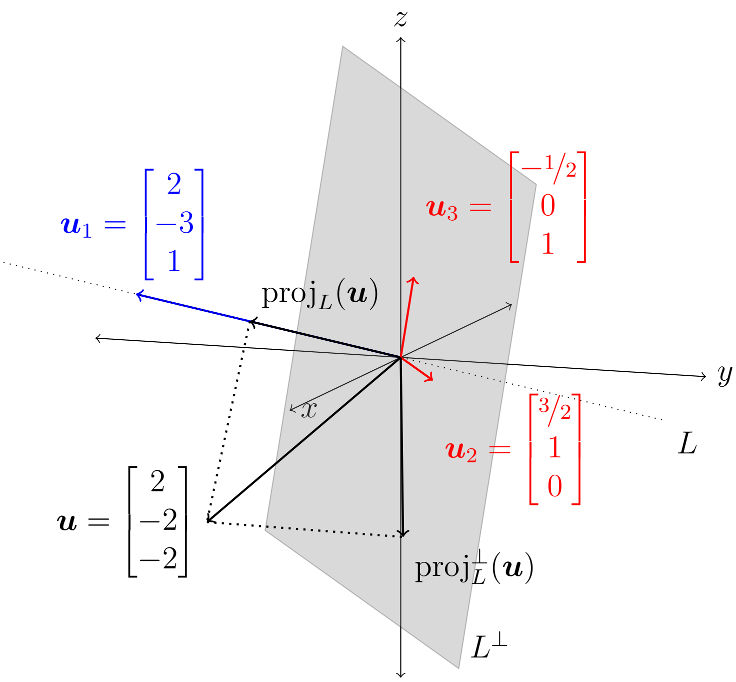
$$\operatorname{rref}\left(\begin{bmatrix}2 & -3 & 1\end{bmatrix}\right) = \begin{bmatrix}1 & -3/2 & 1/2\end{bmatrix}$$

so the special solutions for $\begin{bmatrix}2 & -3 & 1\end{bmatrix}\boldsymbol{x} = \mathbf{0}$ form the canonical basis for $\operatorname{NS}\left(\begin{bmatrix}2 & -3 & 1\end{bmatrix}\right)$, and hence, the canonical basis for

$\mathrm{NS}\left(\begin{bmatrix}2 & -3 & 1\end{bmatrix}\right)$ is

$$\mathcal{B} = \left\{ \begin{bmatrix} 3/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

and so the plane $2x - 3y + z = 0$ is spanned by $\mathcal{B}$.



Problem 128 Let $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$ and let $L = \mathrm{Span}\{\boldsymbol{a}\}$, find the matrices P_L and $P_L^\perp$. Find $\mathrm{proj}_L(\boldsymbol{b})$ and $\mathrm{proj}_L^\perp(\boldsymbol{b})$ for $\boldsymbol{b} = \begin{bmatrix} -2 \\ -3 \\ 0 \end{bmatrix}$ and find the canonical basis for $L^\perp$ from a^T.

All of the above we now generalize to projections into higher dimensional subspaces.

4.1 Projections along a subspace into a subspace

Projection along a subspace $W \subseteq \mathbb{R}^n$ onto a subspace $U \subseteq \mathbb{R}^n$ was introduced in Problem89) . Briefly recall the setup: $\mathbb{R}^n = U \oplus W$, this means $U + W = \mathbb{R}^n$ and $U \cap W = \{\mathbf{0}\}$. The function $\mathrm{proj}_{W,U} : \mathbb{R}^n \to U$ was defined by

$$\mathrm{proj}_{W,U}(\boldsymbol{x}) = \boldsymbol{u} \overset{\mathrm{df}}{\Longleftrightarrow} \boldsymbol{x} = \boldsymbol{u} + \boldsymbol{w} \text{ for some } \boldsymbol{w}$$

This function is linear and hence is given by a matrix P, that is, $\mathrm{proj}_{W,U}(\boldsymbol{x}) = P\boldsymbol{x}$. The projection satisfies $\mathrm{proj}_{W,U}(\boldsymbol{x}) \in U$ and

hence $\text{proj}_{W,U}\left(\text{proj}_{W,U}(\boldsymbol{x})\right) = \text{proj}_{W,U}(\boldsymbol{x})$ so

$$\boxed{\left(P_{W,U}\right)^2 \boldsymbol{x} = P_{W,U}P_{W,U}\boldsymbol{x} = P_{W,U}\boldsymbol{x}}.$$

Some additional properties of this projection matrix are:

- $\text{CS}(P_{W,U}) = U$ and
- $\text{NS}(P_{W,U}) = W$.

Any $n \times n$ matrix satisfying $P^2 = P$ is called a *projection matrix.* The reason for this is explained in the next problem.

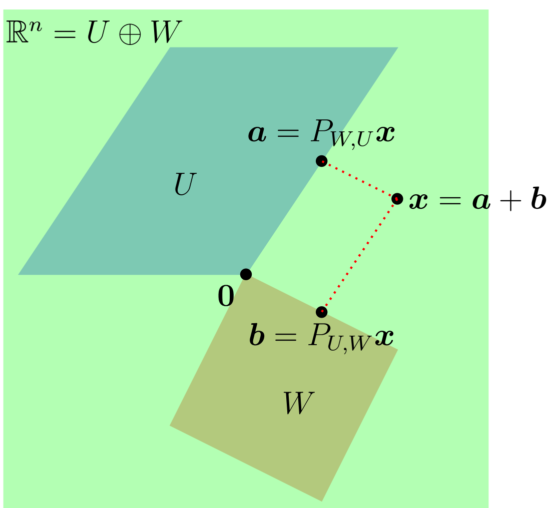
Problem 129 Let $P \in M_{nn}$ be a projection matrix. Show that if $U = \text{CS}(P)$ and $W = \text{NS}(P)$, then

(a) $\mathbb{R}^n = U \oplus W$, that is $\mathbb{R}^n = U + W$ and $U \cap W = \{\mathbf{0}\}$.

(b) P is the matrix for $\text{proj}_{W,U}$.

(c) $I - P$ is a projection matrix and is the matrix for $\text{proj}_{W,U}$.

The picture here is: P is a projection, $W = \mathrm{NS}(P)$, and $U = \mathrm{CS}(P)$:



Given that $\mathbb{R}^n = U \oplus W$ and bases $\mathcal{B}_U = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$, for U, and $\mathcal{B}_W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{(n-r)})\}$, for W. Let A be the $n \times n$ matrix whose columns are the vectors of $\mathcal{B} = \mathcal{B}_U \cup \mathcal{B}_W$. A is invertible, since $\mathcal{B}$ is a basis for $\mathbb{R}^n$. Let $\boldsymbol{x} \in \mathbb{R}^n$ and let

$$\left[\begin{array}{c} \boldsymbol{a} \\ \hdashline \boldsymbol{b} \end{array}\right] = \left[\begin{array}{c} a_1 \\ \vdots \\ a_r \\ \hdashline b_1 \\ \vdots \\ b_{(n-r)} \end{array}\right]$$

be the *unique solution to* $A\boldsymbol{y} = \boldsymbol{x}$, so

$$\left[\begin{array}{c} \boldsymbol{a} \\ \hdashline \boldsymbol{b} \end{array}\right] = A^{-1}\boldsymbol{x}$$

Now $A = \left[\begin{array}{c:c} A_U & A_W \end{array}\right]$ and

$\boldsymbol{x} = AA^{-1}\boldsymbol{x} = \left[\begin{array}{c:c} A_U & A_W \end{array}\right] \begin{bmatrix} \boldsymbol{a} \\ -\boldsymbol{b} \end{bmatrix} = A_U\boldsymbol{a} + A_W\boldsymbol{b}$. Noting that

$$\boldsymbol{a} = \left[\begin{array}{c:c} I_r & O_{r\times(n-r)} \end{array}\right] A^{-1}\boldsymbol{x} \quad \boldsymbol{b} = \left[\begin{array}{c:c} O_{(n-r)\times(r)} & I_{(n-r)} \end{array}\right] A^{-1}\boldsymbol{x}$$

We get

$$\begin{aligned}
\boldsymbol{x} &= A_U\boldsymbol{a} + A_W\boldsymbol{b} \\
&= A \left[\begin{array}{c:c} I_r & O_{r\times(n-r)} \\ \hdashline O_{(n-r)\times(r)} & O_{(n-r)\times(n-r)} \end{array}\right] A^{-1}\boldsymbol{x} + A \left[\begin{array}{c:c} O_{r\times r} & O_{r\times(n-r)} \\ \hdashline O_{(n-r)\times(r)} & I_{(n-r)} \end{array}\right] A^{-1}\boldsymbol{x} \\
&= P_{W,U}\boldsymbol{x} + P_{U,W}\boldsymbol{x}
\end{aligned}$$

So we get

$$\begin{aligned}
P_{W,U} &= A \left[\begin{array}{c:c} I_r & O_{r\times(n-r)} \\ \hdashline O_{(n-r)\times(r)} & O_{(n-r)\times(n-r)} \end{array}\right] A^{-1} \\
P_{U,W} &= A \left[\begin{array}{c:c} O_{r\times r} & O_{r\times(n-r)} \\ \hdashline O_{(n-r)\times(r)} & I_{(n-r)} \end{array}\right] A^{-1} = I - P_{W,U}
\end{aligned}$$

Problem 130 Find $P_{W,U}$ and $P_{U,W}$ for

$$U = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\} \text{ and } W = \text{Span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Find the projection of $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix}$ along W into U.

4.1.1 Orthogonal Projections

The two subspaces U and W, in $\mathbb{R}^n$, are called *orthogonal* iff $\boldsymbol{u} \cdot \boldsymbol{w} = 0$ for all $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in W$. U and W are *orthogonal complements* iff U and W are orthogonal and $U + W = \mathbb{R}^n$.

Problem 131 (a) Are two planes in $\mathbb{R}^3$ ever orthogonal?

(b) Are two planes in $\mathbb{R}^4$ ever orthogonal?

(c) What is the orthogonal complement of a plane in $\mathbb{R}^3$? In $\mathbb{R}^4$?

Suppose W is a subspace of $\mathbb{R}^n$ and $W \oplus W^\perp = \mathbb{R}^n$, define

$$\mathrm{proj}_W = \mathrm{proj}_{W^\perp, W} \qquad \text{and} \qquad \mathrm{proj}_W^\perp = \mathrm{proj}_{W^\perp} = \mathrm{proj}_{W, W^\perp}$$

The matrices in this case are P_W and $P_{W^\perp} = P_W^\perp$.

In such a case we have that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\mathrm{proj}_W(\boldsymbol{y}) \perp \mathrm{proj}_W^\perp(\boldsymbol{x}) = \boldsymbol{x} - \mathrm{proj}_W(\boldsymbol{x})$$

so if P_W is the matrix for proj_W, then

$$(I - P_W)\boldsymbol{x} \cdot P_W \boldsymbol{y} = \boldsymbol{y}^T P_W^T (I - P_W)\boldsymbol{x} = \boldsymbol{y}^T P_W^T \boldsymbol{x} - \boldsymbol{y}^T P_W^T P_W \boldsymbol{x} = 0$$

So $\boldsymbol{y}^T P_W^T \boldsymbol{x} = \boldsymbol{y}^T (P_W^T P_W)\boldsymbol{x}$ for **all** $\boldsymbol{x}, \boldsymbol{y}$ and hence $P_W^T = P_W^T P_W$. So P_W symmetric. The same argument shows P_W is Hermitian in the complex case.

There is a nice converse to this: Call a $n \times n$ matrix an *orthogonal projection matrix* iff

- $PP\boldsymbol{x} = P\boldsymbol{x}$
- P is symmetric (Hermitian.)

Problem 132 Prove that if P is an orthogonal projection matrix and $U = \mathrm{CS}(P)$ and $W = \mathrm{NS}(P)$, then $W = U^\perp$, so P is the matrix of the orthogonal projection proj_U.

Define the distance between a point $\boldsymbol{v}$ and a subspace U as

$$\mathrm{dist}(\boldsymbol{v}, U) = \min\{\mathrm{dist}(\boldsymbol{v}, \boldsymbol{u}) \mid \boldsymbol{u} \in U\}.$$

If $\hat{\boldsymbol{v}} = P\boldsymbol{v}$ where P is an orthogonal projection matrix, then for any other $\boldsymbol{u} \in U$ we have

$$\begin{aligned}
\mathrm{dist}(\boldsymbol{v}, \boldsymbol{u})^2 &= \|\boldsymbol{v} - \boldsymbol{u}\|^2 = \|(\boldsymbol{v} - \hat{\boldsymbol{v}}) + (\hat{\boldsymbol{v}} - \boldsymbol{u})\|^2 \\
&= \|(\boldsymbol{v} - \hat{\boldsymbol{v}})\|^2 + \|(\hat{\boldsymbol{v}} - \boldsymbol{u})\|^2 \quad \text{(Pythagorean Theorem)} \\
&\geq \|(\boldsymbol{v} - \hat{\boldsymbol{v}})\|^2 = \mathrm{dist}(\boldsymbol{v}, \hat{\boldsymbol{v}})^2
\end{aligned}$$

The Pythagorean Theorem applies since $\boldsymbol{v} - \hat{\boldsymbol{v}} \perp \hat{\boldsymbol{v}} - \boldsymbol{u}$.

So $\mathrm{dist}(\boldsymbol{v}, \hat{\boldsymbol{v}}) \leq \mathrm{dist}(\boldsymbol{v}, \boldsymbol{u})$ and thus

$$\mathrm{dist}(\boldsymbol{v}, U) = \mathrm{dist}(\boldsymbol{v}, P\boldsymbol{v}).$$

Notice $\mathrm{dist}(\boldsymbol{v}, U)$ is always > 0 for $\boldsymbol{v} \notin U$, this was not clear from the initial definition.

As a result we have

$$\mathrm{proj}_U(\boldsymbol{v}) = \text{ the point in } U \text{ closest to } \boldsymbol{v}.$$

Now we address finding P_U and hence $\mathrm{proj}_U(\boldsymbol{v})$. The following problem is just Problem 114. The problem is repeated here in slightly different language.

Problem 133 Given an $m \times n$ matrix A with $\mathrm{rank}(A) = r$ show that the $n \times n$ matrix $A^T A$ has $\mathrm{rank}(A^T A) = r$. Hint: Show

$\mathrm{NS}(A) = \mathrm{NS}(A^TA)$. Use $A^TA\boldsymbol{x} = \mathbf{0} \Rightarrow \boldsymbol{x}^TA^TA\boldsymbol{x} = 0$ and interpret as the dot product.

It follows from the preceding that if A has n linearly independent columns, then A^TA is invertible.

Given a basis $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ for U let $A = \left[\, \boldsymbol{u}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{u}_k \,\right]$ be the $m \times k$ matrix whose i^{th} column is just the i^{th} member of the basis, then A^TA is invertible. We wish to find matrix P so that

- $A^T(\boldsymbol{u} - P\boldsymbol{u}) = \mathbf{0}$ and
- $P\boldsymbol{u} = A\boldsymbol{y}$ for some $\boldsymbol{y}$

so we are looking for

$$A^T\boldsymbol{u} = A^TA\boldsymbol{y}.$$

Since A^TA is invertible we get

$$\boldsymbol{y} = (A^TA)^{-1}A^T\boldsymbol{u}$$

and

$$P\boldsymbol{u} = A\boldsymbol{y} = A(A^TA)^{-1}A^T\boldsymbol{u}.$$

Since this holds for all $\boldsymbol{u}$,

$$\boxed{P = A(A^TA)^{-1}A^T}.$$

Problem 134 Check that $P = A(A^TA)^{-1}A^T$ is an orthogonal projection matrix, that is $PP\boldsymbol{x} = \boldsymbol{x}$ and $P^T = P$.

Remark This gives a second proof that $P = A(A^TA)^{-1}A^T = P_U$. This is because we know $P = P_{\mathrm{CS}(P)}$ so all we need is that $\mathrm{CS}(P) = U$. Clearly, $\mathrm{CS}(P) \subset \mathrm{CS}(A) = U$ and for any $\boldsymbol{v} \in \mathrm{CS}(A)$, say $\boldsymbol{v} = A\boldsymbol{u}$ we have

$$P\boldsymbol{v} = A(A^TA)^{-1}A^T(A\boldsymbol{u}) = A[(A^TA)^{-1}(A^TA)]\boldsymbol{u} = AI\boldsymbol{u} = A\boldsymbol{u} = \boldsymbol{v}$$

So $\boldsymbol{v} \in \mathrm{CS}(P)$. ♢

Problem 135 Find the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ into

$\mathrm{Span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 4 \end{bmatrix} \right\}.$

We can finally prove that $W^\perp$ is the a complement of W, that is, $\mathbb{R}^n = W \oplus W^\perp$.

Problem 136 For W a subspace of $\mathbb{R}^n$ show that $\mathbb{R}^n = W \oplus W^\perp$, that is, show that

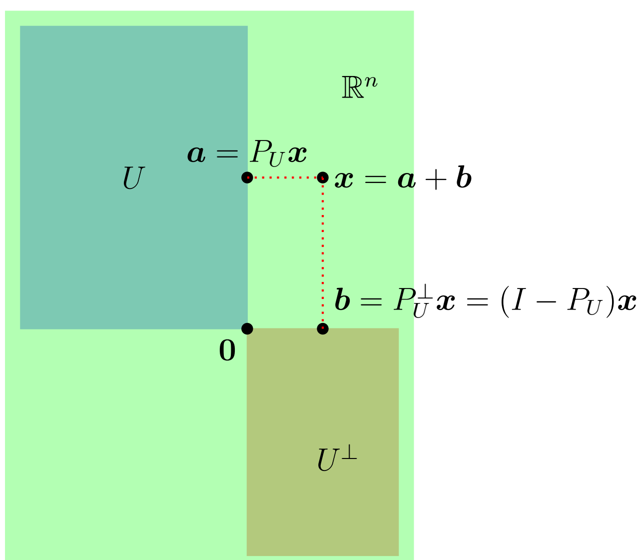
- $W \cap W^\perp = \{\mathbf{0}\}$ and
- $\mathbb{R}^n = W + W^\perp$

So W and $W^\perp$ are orthogonal complements.

To summarize: Given $W = \mathrm{Span}(\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\})$ where the $\boldsymbol{w}_i$'s are linearly independent, letting $A = \begin{bmatrix} \boldsymbol{w}_1 & | & \boldsymbol{w}_2 & | & \cdots & | & \boldsymbol{w}_k \end{bmatrix}$ we have

$$\begin{aligned} \mathrm{proj}_W(\boldsymbol{u}) &= A(A^TA)^{-1}A^T\boldsymbol{u} \\ \mathrm{proj}_W^\perp(\boldsymbol{u}) &= \boldsymbol{u} - A(A^TA)^{-1}A^T\boldsymbol{u} \\ \mathrm{dist}(\boldsymbol{u}, W) &= \|\boldsymbol{u} - A(A^TA)^{-1}A^T\boldsymbol{u}\| \end{aligned}$$

The picture in this general case can be represented by



Problem 137 Find $\mathrm{proj}_W(\boldsymbol{u})$, $\mathrm{proj}_W^\perp(\boldsymbol{u})$, and $\mathrm{dist}(\boldsymbol{u}, W)$ where $W = \mathrm{Span}(\{(1,2,3,4), (-1,2,0,1), (1,1,-2,2)\})$ and $\boldsymbol{u} = (-2,2,1,0)$.

4.2 Least Squares solution to $A\mathrm{x} = \mathbf{b}$ for A of full column rank

We look at finding best approximate solutions for m linear equations in $n < m$ unknowns, $A\boldsymbol{x} = \boldsymbol{b}$ for $A \in M_{mn}$. Typically there is no solution in this case.

If the system $A\boldsymbol{x} = \boldsymbol{b}$ fails to have a solution, one can still look for the *best approximate solution*, namely that $\boldsymbol{x}$ so that $\operatorname{dist}(A\boldsymbol{x}, \boldsymbol{b})$ is minimized. This just amounts to finding a point in $\mathrm{CS}(A)$ so that $\operatorname{dist}(\boldsymbol{b}, \mathrm{CS}(A))$ is minimized.

In the case that A has full column rank we have $\hat{\boldsymbol{b}} = \operatorname{proj}_{\mathrm{CS}(A)}(\boldsymbol{b})$, and we look for those $\hat{\boldsymbol{x}}$ such that $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$. Since the columns of A are linearly independent, $\hat{\boldsymbol{x}}$ is unique and we have

$$A\hat{\boldsymbol{x}} = A(A^TA)^{-1}A^T\boldsymbol{b}$$

multiplying by A^T on both sides gives

$$(A^TA)\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \qquad (\star)$$

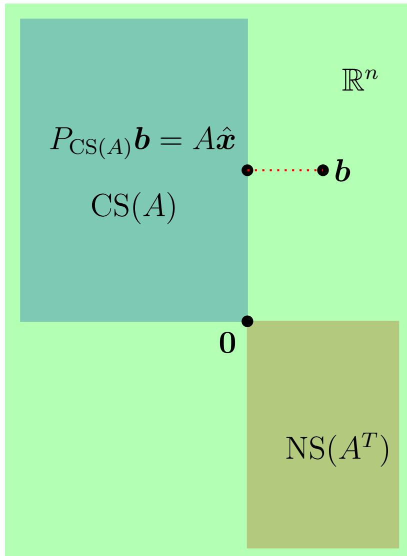
since A^TA is invertible we get

$$\hat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b} \qquad (\star)$$

This $\hat{\boldsymbol{x}}$ is called the *least squares solution to* $A\boldsymbol{x} = \boldsymbol{b}$, this is the unique $\hat{\boldsymbol{x}}$ that minimizes $\operatorname{dist}(A\hat{\boldsymbol{x}}, \boldsymbol{b}) = \|A\hat{\boldsymbol{x}} - \boldsymbol{b}\|$. This is the least

squares solution because you are minimizing $\sqrt{\sum_{i=1}^{k}\big((A\hat{\boldsymbol{x}})_i - b_i\big)^2}$, which is equivalent to minimizing $\sum_{i=1}^{k}\big((A\hat{\boldsymbol{x}})_i - b_i\big)^2$.

Recall this very same matrix was used back on page 119 to solve $A\boldsymbol{x} = \boldsymbol{b}$ in the case that A had linearly independent columns **and** $\boldsymbol{b}$ was in the column space of A



Find $\hat{\boldsymbol{x}}$ that minimizes $\|A\hat{\boldsymbol{x}} - \boldsymbol{b}\|$.

Problem 138 Find the least squares solution to

(a) $$\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 4 \\ 0 & 1 \\ 7 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 4 \\ 3 \\ 1 \\ 2 \end{bmatrix}$$

Hint: Recall all you need to do is solve $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$, as long as the columns of A are linearly independent you will get a unique solution $\hat{\boldsymbol{x}}$.

(b) $$\begin{bmatrix} 1 & 2 & 2 \\ 3 & -1 & 1 \\ 2 & 4 & 4 \\ 0 & 1 & 0 \\ 7 & 4 & 6 \\ 5 & 6 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 4 \\ 3 \\ 1 \\ 2 \end{bmatrix}$$

4.3 Fitting polynomials to data

Fitting polynomials, e.g., lines, parabolas, etc. to data is a typical application of least squares solutions. Suppose we want to find a line that "best fits" some data (x_i, y_i) for $i = 1, \ldots, N$. We look for a, b so that $y_i = ax_i + b$, of course there is no reason such a line should exist so we look for the line that minimizes, *the total error*,

i.e., the sum of all of the errors $e_i = |y_i - (ax_i + b)|$. It turns out that minimizing the sum of the squares of these errors an easier question. Rephraing the question as a matrix equation we arte trying to solve

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}}_{A} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

This need not have a solution, but we can find the least squares solution by solving:

$$A^T A \begin{bmatrix} b \\ a \end{bmatrix} = A^T \boldsymbol{y} \qquad (\star)$$

This will minimize the error

$$\left\| A \begin{bmatrix} b \\ a \end{bmatrix} - \boldsymbol{y} \right\| = \left(\sum_{i=1}^{N} \left((b + ax_i) - y_i \right)^2 \right)^{1/2} = \left(\sum_{i=1}^{N} e_i^2 \right)^{1/2}$$

(least squares error)

Problem 139 Write out $(\star)$ to get the typical formulation of the regresion line from statistics:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

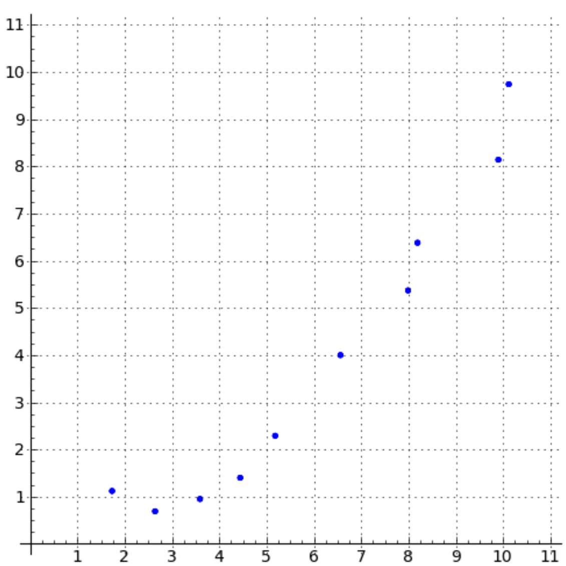
$$\begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

now solve for a and b.

Problem 140 Consider the data

$$\begin{aligned} D = &\{(1.73, 1.12), (2.64, 0.69), (3.59, 0.95), (4.44, 1.40), (5.18, 2.29), \\ &(6.56, 4.00), (7.99, 5.37), (8.19, 6.38), (9.90, 8.14), (10.12, 9.74)\} \end{aligned}$$

(a) Find the straight line that best fits the data. Sketch the line.



(b) Generalize the above method to find the quadratic that best fits the data. Sketch the quadratic approximation on the same graph you used in (a).

(c) Which approximation fits the data best? (Which has least sum of squares error?)

4.4 Gramm-Schmidt, orthonormal bases, QR decomposition

A set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is *orthogonal* iff $\boldsymbol{v}_i \perp \boldsymbol{v}_j$ for $i \neq j$ (and $\boldsymbol{v}_i \neq \mathbf{0}$ for all i).

Problem 141 Let $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be an orthogonal set of vectors. show

(a) If $\boldsymbol{v} = \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i \in \operatorname{Span} \mathcal{B}$, then

$$\alpha_i = \frac{\boldsymbol{v} \cdot \boldsymbol{v}_i}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i} = [(\boldsymbol{v}_i^T \boldsymbol{v}_i)^{-1} \boldsymbol{v}_i^T] \boldsymbol{v}$$

(b) $\mathcal{B}$ is linearly independent.

Problem 142 Show that it $\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is an orthogonal basis for U, a subspace of $\mathbb{R}^n$, then

$$\begin{aligned}
\mathrm{proj}_U(\boldsymbol{u}) &= \sum_{i=1}^{k} \mathrm{proj}_{\boldsymbol{u}_i}(\boldsymbol{u}) \\
&= \sum_{i=1}^{k} \left(\frac{\boldsymbol{u} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i} \right) \boldsymbol{u}_i \\
&= \sum_{i=1}^{k} \left[\boldsymbol{u}_i (\boldsymbol{u}_i^T \boldsymbol{u}_i)^{-1} \boldsymbol{u}_i^T \right] \boldsymbol{u}
\end{aligned}$$

A useful and obvious fact is that if U and W are orthogonal subspaces of $\mathbb{R}^n$ with $\mathcal{B}_U$ and $\mathcal{B}_W$ orthogonal bases for U and W respectively, then $\mathcal{B}_U \cup \mathcal{B}_W$ is an orthogonal basis for $U \oplus W$.

A set of vectors is *orthonormal* if it is orthogonal and each vector is of unit length. A matrix A is *orthogonal* provided the columns of A are an orthonormal set of vectors.[16]

Problem 143 Show that the following are equivalent for an $m \times n$ orthogonal matrix. (Note A is of full column rank so $m \geq n$.)

(1) A is orthogonal.

(2) $A^T A = I_n$.

(3) $(A\boldsymbol{x}) \cdot (A\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$.

(4) $\|A\boldsymbol{x}\| = \|\boldsymbol{x}\|$ so $A : \mathbb{R}^n \to \mathbb{R}^m$ preserves lengths.

(5) $\operatorname{dist}(A\boldsymbol{x}_1, A\boldsymbol{x}_2) = \operatorname{dist}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ so $A : \mathbb{R}^n \to \mathbb{R}^m$ preserves distances.

[16] There is no particular reason for not having two distinct notions, that of an orthogonal matrix and that of an orthonormal matrix, this is just a matter of history.

If in the previous problem A is square, then A is called *unitary* and we can add to the list of equivalences $A^T = A^{-1}$.

Problem 144 Let $\boldsymbol{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$. Show

(a) For any $\boldsymbol{u} \in \operatorname{Span} B$, $\boldsymbol{u} = \sum_{i=1}^{k} (\boldsymbol{u} \cdot \boldsymbol{u}_i)\boldsymbol{u}_i$.

(b) Let $A = \left[\, \boldsymbol{u}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{u}_k \,\right]$, then if $\boldsymbol{x} \in \mathrm{CS}(A)$ and $\boldsymbol{a} = A^T\boldsymbol{x}$, it follows that $\boldsymbol{x} = A\boldsymbol{a}$ so we have

$$AA^T\boldsymbol{x} = \boldsymbol{x} \text{ for all } \boldsymbol{x} \in \mathrm{CS}(A)$$
$$A^TA\boldsymbol{a} = \boldsymbol{a} \text{ for all } \boldsymbol{a} \in \mathbb{R}^k$$

Suppose we are given a linear independent set of vectors $\mathcal{U}=\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_k\}$ and want to convert this into an orthonormal $\mathcal{Q}$ with $\operatorname{Span}\mathcal{U}=\operatorname{Span}\mathcal{Q}$. A method to do this is *Gramm-Schmidt*

Step 1 ("The hard part.") First find an orthogonal set $\mathcal{V}$ with the same span as $\mathcal{U}$. To do this proceed as follows:

$$
\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{u}_1\\
\boldsymbol{v}_2 &= \boldsymbol{u}_2 - \left(\frac{\boldsymbol{u}_2\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\right)\boldsymbol{v}_1 = \boldsymbol{u}_2 - \operatorname{proj}_{\boldsymbol{v}_1}(\boldsymbol{u}_2)\\
&= \boldsymbol{u}_2 - \operatorname{proj}_{\operatorname{Span}\{\boldsymbol{u}_1\}}(\boldsymbol{u}_2) = \operatorname{proj}^{\perp}_{\operatorname{Span}\{\boldsymbol{u}_1\}}(\boldsymbol{u}_2)\\
\boldsymbol{v}_3 &= \boldsymbol{u}_3 - \left(\frac{\boldsymbol{u}_3\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\right)\boldsymbol{v}_1 - \left(\frac{\boldsymbol{u}_3\cdot\boldsymbol{v}_2}{\boldsymbol{v}_2\cdot\boldsymbol{v}_2}\right)\boldsymbol{v}_2 = \boldsymbol{u}_3 - \operatorname{proj}_{\boldsymbol{v}_1}(\boldsymbol{u}_3) - \operatorname{proj}_{\boldsymbol{v}_2}(\boldsymbol{u}_3)\\
&= \boldsymbol{u}_3 - \operatorname{proj}_{\operatorname{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2\}}(\boldsymbol{u}_3) = \operatorname{proj}^{\perp}_{\operatorname{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2\}}(\boldsymbol{u}_3)\\
&\vdots\\
\boldsymbol{v}_{i+1} &= \boldsymbol{u}_{i+1} - \sum_{j=1}^{i}\left(\frac{\boldsymbol{u}_{i+1}\cdot\boldsymbol{v}_j}{\boldsymbol{v}_j\cdot\boldsymbol{v}_j}\right)\boldsymbol{v}_j = \boldsymbol{u}_{i+1} - \sum_{j=1}^{i}\operatorname{proj}_{\boldsymbol{v}_j}(\boldsymbol{u}_{i+1})\\
&= \boldsymbol{u}_{i+1} - \operatorname{proj}_{\operatorname{Span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_i\}}(\boldsymbol{u}_{i+1}) = \operatorname{proj}^{\perp}_{\operatorname{Span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_i\}}(\boldsymbol{u}_{i+1})
\end{aligned}
$$

Letting

$$A_i = \left[\, \boldsymbol{u}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{u}_i \,\right]$$

we have

$$\boldsymbol{v}_{i+1} = \left[I - A_i(A_i^TA_i)^{-1}A_i^T\right]\boldsymbol{u}_{i+1}.$$

That $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_i\}$ is orthogonal and $\operatorname{Span}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_i\}=\operatorname{Span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_i\}$ is evident.

Step 2: Let $\boldsymbol{q}_i = \dfrac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|}$.

For one reason or another, it seems $\mathcal{Q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_k\}$ is often used to denote an orthonormal set of vectors, the associated orthogonal matrix is $Q = \left[\boldsymbol{q}_1 \vdots \cdots \vdots \boldsymbol{q}_k\right]$.

Once we have $\mathcal{Q}$ notice that since $\boldsymbol{u}_i \in \operatorname{Span}\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_i\}$ we have (see also Problem 144)

$$\operatorname{col}_i(A) = \boldsymbol{u}_i = QQ^T\boldsymbol{u}_i = Q\begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{u}_i \\ \boldsymbol{q}_2^T\boldsymbol{u}_i \\ \vdots \\ \boldsymbol{q}_i^T\boldsymbol{u}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Q\operatorname{col}_i(Q^TA)$$

so $A = Q(Q^TA)$ and $R = Q^TA$ is the $k \times k$ upper triangular matrix with

$$\operatorname{col}_i(R) = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{u}_i \\ \boldsymbol{q}_2^T\boldsymbol{u}_i \\ \vdots \\ \boldsymbol{q}_i^T\boldsymbol{u}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

R has non-zero entries on the diagonal so R is invertible.

$$\boxed{A = QR \text{ is a } QR \textit{ decomposition}}.$$

Problem 145 Use Gramm-Schmidt to find an orthonormal basis for $\mathrm{CS}(A)$ and then find the associated QR-decomposition for A.

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

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

Given the QR decomposition of $A \in M_{mn}$ $(m > n)$ the least squares problem

$$A^T A\hat{\boldsymbol{x}} = A^T \boldsymbol{b}$$

becomes

$$R^T Q^T QR\hat{\boldsymbol{x}} = R^T I_n R\hat{\boldsymbol{x}} = R^T R\hat{\boldsymbol{x}} = R^T Q^T \boldsymbol{b}$$

which gives the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ to be $\hat{\boldsymbol{x}}$ such that:

$$\boxed{R\hat{\boldsymbol{x}} = Q^T \boldsymbol{b}}$$

since R, and hence R^T, is invertible. This last equation is particularly easy to solve by substitution since R is upper triangular.

Problem 146 For the matrix in part (b) of Problem 145 check that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution and then find the least squares solution to $A\hat{\boldsymbol{x}} = \boldsymbol{b}$ by solving $R\hat{\boldsymbol{x}} = Q^T \boldsymbol{b}$ for $\boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$. In addition, find the least squares error, $\|A\hat{\boldsymbol{x}} - \boldsymbol{b}\|$.

5 Linear mappings and inner product spaces

For V and W vector spaces a map $L : V \to W$ is *linear* iff L preserves linear compositions, that is,

$$L(\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k) = \alpha_1 L(\boldsymbol{v}_1) + \cdots + \alpha_k L(\boldsymbol{v}_k)$$

To check that L is linear it suffices to check that L preserves scalar multiplication and vector addition, that is,

- $L(\alpha \boldsymbol{v}) = \alpha L(\boldsymbol{v})$
- and $L(\boldsymbol{v}_1 + \boldsymbol{v}_2) = L(\boldsymbol{v}_1) + L(\boldsymbol{v}_2)$.

The generic example of a linear mapping is given by matrix multiplication, when A is an $m \times n$ matrix, then $L_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by $L_A(\boldsymbol{x}) = A\boldsymbol{x}$ is a linear mapping since matrix multiplication is linear, that is, $A(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha(A\boldsymbol{x}) + \beta(A\boldsymbol{y})$.

Problem 147 Let $L : V \to W$ be linear prove the following:

(a) $\mathrm{Ker}(L) = \{\boldsymbol{v} \mid L(\boldsymbol{v}) = \boldsymbol{0}\}$ is a subspace of V. (This is the analogue of $\mathrm{NS}(A)$.)

(b) $\mathrm{Img}(L) = \{L(\boldsymbol{v}) \mid \boldsymbol{v} \in V\}$ is a subspace of W. (This is analogous to $\mathrm{CS}(A)$.)

Problem 148 Define $P : V \to V$ to be a projection iff $P^2 = P$, that is $P(P(\boldsymbol{v})) = P(\boldsymbol{v})$. Let $N = \mathrm{Ker}(L)$ and let $U = \mathrm{Img}(L)$. Show that $U \oplus \mathrm{Ker}(N) = V$. (Hint: $\boldsymbol{v} = P(\boldsymbol{v}) + (\mathrm{id} - P)(\boldsymbol{v})$)

In general given $V = V_1 \oplus V_2$ and $\boldsymbol{v} \in V$ there are unique $\boldsymbol{v}_1 \in V_1$ and $\boldsymbol{v}_2 \in V_2$. Define $P_{V_1,V_2}(\boldsymbol{v}) = \boldsymbol{v}_2$ and $P_{V_2,V_1}(\boldsymbol{v}) = \boldsymbol{v}_1$. These two maps are linear and satisfy

$$
\begin{aligned}
&P^2_{V_1,V_2} = P_{V_1,V_2} \quad \mathrm{Img}(P_{V_1,V_2}) = V_2 \quad \mathrm{Ker}(P_{V_1,V_2}) = V_1 \\
&P^2_{V_2,V_1} = P_{V_2,V_1} \quad \mathrm{Img}(P_{V_2,V_1}) = V_1 \quad \mathrm{Ker}(P_{V_2,V_1}) = V_2 \\
&P_{V_1,V_2} = \mathrm{id} - P_{V_2,V_1} \quad P_{V_2,V_1} = \mathrm{id} - P_{V_1,V_2}.
\end{aligned}
$$

So abstractly $P^2 = P$ characterizes projections and for such a P, $P = P_{\mathrm{Ker}(P),\mathrm{Img}(P)}$.

To generalize the notion of orthogonal projection we need to generalize the notion of adjoint - hence self-adjoint - to linear operators. This will be done when discussing inner product spaces below.

The next problem generalizes the fact that $\dim(\mathrm{RS}(A)) = \dim(\mathrm{CS}(A))$. The point is $\mathbb{R}^n = \mathrm{RS}(A) \oplus \mathrm{NS}(A)$ for $A \in M_{mn}$.

Problem 149 Show that if $L : V \to W$ and $V = U \oplus \mathrm{Ker}(L)$, then $L|_U : U \to \mathrm{Img}(L)$ is one-to-one and onto. Argue that if $\mathcal{B}$ is a basis for U, then $L[\mathcal{B}] = \{L(\boldsymbol{v}) \,|\, \boldsymbol{v} \in \mathcal{B}\}$ is a basis for $\mathrm{Img}(L)$.

The preceding brings up the issue of *invariant subspaces.* If $L \in \mathcal{L}(V)$ is a linear operator on V and U is a subspace of V, then $L|_U \in \mathcal{L}(U)$ if $L : U \to U$. When this happens call U an *invariant subspace of* L. If $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$ and each V_i is an invariant subspace of L, then L can be fully understood by what L does on each of the V_i's.

For example a projection map $P : V \to V$ has two invariant subspaces, namely $\mathrm{Img}(P)$ and $\mathrm{Ker}(P)$ and $V = \mathrm{Img}(P) \oplus \mathrm{Ker}(P)$. $P|_{\mathrm{Img}(P)} = \mathrm{id}$ while $P|_{\mathrm{Ker}(P)} = \mathbf{0}|_{\mathrm{Ker}(P)}$ so that $P = \mathrm{id}\,|_{\mathrm{Img}(P)} \oplus \mathbf{0}|_{\mathrm{Ker}(P)} : \mathrm{Img}(P) \oplus \mathrm{Ker}(P) \to \mathrm{Img}(P) \oplus \mathrm{Ker}(P)$. In this way a decomposition of V into L-invariant subspaces induces a decomposition of L itself into “simpler” operators on “simpler” spaces.

This idea is what underlies the study of eigenvalues, eigenvectors, eigenspaces, and eigendecomposition below.

5.1 The matrix of a transformation

If V and W are vector spaces with $\mathcal{B}_V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $\mathcal{B}_W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ bases for V and W respectively and $L : V \to W$ is linear, then L is completely determined by how L moves the basis elements from $\mathcal{B}_V$.

Problem 150 Verify the last claim.

Write $L(\boldsymbol{v}_j)$ as its, unique, linear combination of elements of $\mathcal{B}_W$, so $L(\boldsymbol{v}_j) = \sum_{i=1}^m a_{ij}\boldsymbol{w}_i$ where $L(\boldsymbol{v}_j)_{\mathcal{B}_W} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$. For $\boldsymbol{v} \in V$, say

$\boldsymbol{v}_{\mathcal{B}_V} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$, we have:

$$\begin{aligned} L(\boldsymbol{v}) &= \sum_{j=1}^{n} \alpha_j L(\boldsymbol{v}_j) = \sum_{j=1}^{n} \alpha_j \left(\sum_{i=1}^{m} a_{ij} \boldsymbol{w}_i \right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j a_{ij} \right) \boldsymbol{w}_i \\ &= \sum_{i=1}^{m} \left(\left[L(\boldsymbol{v}_1)_{\mathcal{B}_W} \vdots L(\boldsymbol{v}_2)_{\mathcal{B}_W} \vdots \cdots \vdots L(\boldsymbol{v}_n)_{\mathcal{B}_W} \right] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \right) \boldsymbol{w}_i \\ &= \left(A(\boldsymbol{v}_{\mathcal{B}_V})\right)^{\mathcal{B}_W} \end{aligned}$$

where

$$A = [L]_{\mathcal{B}_W, \mathcal{B}_V} = \left[L(\boldsymbol{v}_1)_{\mathcal{B}_W} \vdots L(\boldsymbol{v}_2)_{\mathcal{B}_W} \vdots \cdots \vdots L(\boldsymbol{v}_n)_{\mathcal{B}_W} \right]$$

So

$$L(\boldsymbol{v}) = \left([L]_{\mathcal{B}_W, \mathcal{B}_V} \boldsymbol{v}_{\mathcal{B}_V}\right)^{\mathcal{B}_W}$$

Written another way this is

$$L(\boldsymbol{v})_{\mathcal{B}_W} = [L]_{\mathcal{B}_W, \mathcal{B}_V} \boldsymbol{v}_{\mathcal{B}_V}$$

Here $[L]_{\mathcal{B}_W, \mathcal{B}_V}$ is the matrix that takes a $\mathcal{B}_V$ component representation for $\boldsymbol{v} \in V$ and outputs a $\mathcal{B}_W$ component representation for $L(\boldsymbol{v}) \in W$.

Remark If $L : \mathbb{R}^n \to \mathbb{R}^m$ and the standard bases, $\mathcal{E}^n$ and $\mathcal{E}^m$, are used on both sides, then $[L]_{\mathcal{E}^n, \mathcal{E}^m}$ is just denoted $[L]$.

In particular, if A is an $m \times n$ matrix and $L_A : \mathbb{R}^n \to \mathbb{R}^m$ is, as usual, given by $L_A(\boldsymbol{x}) = A\boldsymbol{x}$, then $[L_A] = A$. This is because

$$\begin{aligned} [L_A]_{\mathcal{E}^m, \mathcal{E}^n} &= \left[L_A(\boldsymbol{e}_1^n)_{\mathcal{E}^m} \vdots \cdots \vdots L_A(\boldsymbol{e}_n^n)_{\mathcal{E}^m} \right] = \left[(A\boldsymbol{e}_1^n)_{\mathcal{E}^m} \vdots \cdots \vdots (A\boldsymbol{e}_n^n)_{\mathcal{E}^m} \right] \\ &= \left[\mathrm{col}_1(A)_{\mathcal{E}^m} \vdots \cdots \vdots \mathrm{col}_n(A)_{\mathcal{E}^m} \right] \\ &= \left[\mathrm{col}_1(A) \vdots \cdots \vdots \mathrm{col}_n(A) \right] = A. \end{aligned}$$

$\diamondsuit$

In this context a change of basis matrix on $\mathbb{R}^n$, changing from $\mathcal{B}$ coordinates to $\mathcal{C}$ coordinates is $[\mathrm{id}]_{\mathcal{B},\mathcal{C}}$. some things to note are

- The change of basis from $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ coordinates to standard coordinates is

$$[\mathrm{id}]_{\mathcal{E},\mathcal{B}} = \left[\boldsymbol{v}_{1\mathcal{E}} \vdots \cdots \vdots \boldsymbol{v}_{n\mathcal{E}} \right] = \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_n \right] = B.$$

  As before the matrix B associated to a basis $\mathcal{B}$ is just the matrix whose columns are the vectors of $\mathcal{B}$ written in terms of the standard coordinates.

- The change of basis from standard to $\mathcal{B}$ is

$$[\mathrm{id}]_{\mathcal{B},\mathcal{E}} = \left[(\boldsymbol{e}_1^n)_{\mathcal{B}} \,\vdots\, (\boldsymbol{e}_2^n)_{\mathcal{B}} \,\vdots\, \cdots \,\vdots\, (\boldsymbol{e}_n^n)_{\mathcal{B}} \right] = ([\mathrm{id}]_{\mathcal{E},\mathcal{B}})^{-1} = B^{-1}.$$

- The change of basis from $\mathcal{B}$ to $\mathcal{C}$ can be obtained by going from $\mathcal{B}$ to standard and then from standard to $\mathcal{C}$ so

$$[\mathrm{id}]_{\mathcal{C},\mathcal{B}} = [\mathrm{id}]_{\mathcal{C},\mathcal{E}}[\mathrm{id}]_{\mathcal{E},\mathcal{B}} = ([\mathrm{id}]_{\mathcal{C},\mathcal{E}})^{-1}[\mathrm{id}]_{\mathcal{E},\mathcal{B}} = C^{-1}B$$

The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$ and is itself a vector space, essentially the same as M_{mn} if $\dim(V) = n$ and $\dim(W) = m$. An important special case is when $V = W$, that is when $L : V \to V$ is linear. In this case call L a *linear operator on* V. The vector space of all linear operators is denoted $\mathcal{L}(V)$. For $\mathcal{B}$ a basis for V and $L \in \mathcal{L}(V)$ write $[L]_{\mathcal{B}}$ for $[L]_{\mathcal{B},\mathcal{B}}$.

Problem 151 Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping projecting points in $\mathbb{R}^3$ to the closest point in $W = \mathrm{Span}\left(\left\{\begin{bmatrix}1\\0\\-2\end{bmatrix}, \begin{bmatrix}1\\-2\\0\end{bmatrix}\right\}\right)$.

Let $\mathcal{B} = \left\{\begin{bmatrix}1\\0\\-2\end{bmatrix}, \begin{bmatrix}1\\-2\\0\end{bmatrix}, \begin{bmatrix}2\\1\\1\end{bmatrix}\right\}$. Noting that

$$L\left(\begin{bmatrix}1\\0\\-2\end{bmatrix}\right) = \begin{bmatrix}1\\0\\-2\end{bmatrix} \quad L\left(\begin{bmatrix}1\\-2\\0\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\0\end{bmatrix} \quad L\left(\begin{bmatrix}2\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

find the following:

(a) $[L]_{\mathcal{B}}$.

(b) $[L]_{\mathcal{E},\mathcal{B}}$.

(c) $[L]_{\mathcal{B},\mathcal{E}}$.

(d) $[L]$.

Two square matrices A and B are *similar* iff there is an invertible matrix P so that $A = PBP^{-1}$.

Problem 152 Show that $n \times n$ matrices A and B are similar iff there is a choice of basis $\mathcal{D}$ for $\mathbb{R}^n$ so that

$$[L_A]_{\mathcal{D}} = B$$

So two matrices are similar if they both represent the same linear

transformation with respect to an appropriate choice of basis.

Problem 153 Let $\boldsymbol{c} = (c_1, \ldots, c_n)$ with $c_i \neq c_j$ for $i \neq j$. Let

$$p_i^{\boldsymbol{c}}(x) = \prod_{\substack{j=1,\ldots,n \\ j \neq i}} \tfrac{(x-c_j)}{(c_i-c_j)}.$$

(a) Show that $\mathcal{B}^{\boldsymbol{c}} = \{p_i^{\boldsymbol{c}}(x) \,|\, i = 1, \ldots, n\}$ is a basis for $P_{n-1}[x]$.

(b) Given an arbitrary $q(x)$ in $P_{n-1}[x]$ show that $q_{\mathcal{B}^{\boldsymbol{c}}} = \begin{bmatrix} q(c_1) \\ q(c_2) \\ \vdots \\ q(c_n) \end{bmatrix}$.

That is $q(x) = \sum_{i=1}^{n} q(c_i) p_i^{\boldsymbol{c}}(x)$.

(c) For $\boldsymbol{c} = (1, 2, 3, 4)$ find the basis $\mathcal{B}^{\boldsymbol{c}}$ for $P_3[x]$.

(d) Find $q_{\mathcal{B}^{\mathbf{c}}}$ where $q(1) = 1$, $q(2) = 3$, $q(3) = -4$, and $q(4) = 0$ and then find $q = (q_{\mathcal{B}^{\mathbf{c}}})^{\mathcal{B}^{\mathbf{c}}}$.

(e) Find the change of bases matrices from $\mathcal{B}^{\mathbf{c}}$ to the standard basis for $P_3[x]$ and from the standard basis back to $\mathcal{B}^{\mathbf{c}}$.

(f) For $\boldsymbol{d} = (0, 1, 2)$ find the basis $\mathcal{B}^{\boldsymbol{d}}$ for $P_2[x]$

(g) Find $[\frac{d}{dx}]_{\mathcal{B}^d,\mathcal{B}^c}$ where $\frac{d}{dx} : P_3[x] \to P_2[x]$ is just differentiation.

Problem 154 The operation $\cdot^T : M_{33} \to M_{33}$ is linear, find the matrix for this transformation wrt the standard basis for M_{33}, so E_{ij} is the 3×3 matrix satisfying $E_{ij}(rs) = \begin{cases} 1 & (i,j) = (r,s) \\ 0 & \text{otherwise} \end{cases}$.
Take this basis to be ordered lexicographically, i.e.,
$\mathcal{E} = \{E_{11}, E_{12}, E_{13}, E_{21}, \ldots, E_{32}, E_{33}\}$ so $(E_{11})_{\mathcal{E}} = \boldsymbol{e}_1^9$, $(E_{12})_{\mathcal{E}} = \boldsymbol{e}_2^9$, etc.

Problem 155 Show that if $L : V \to W$ is linear and invertible, i.e., one-to-one and onto, then the inverse $L^{-1} : W \to V$ is linear. (Give a proof that works even when V, and hence W, are infinite dimensional.)

Problem 156 What can you say about the form of $[L]_{\mathcal{B}}$ if $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, where each V_i is an L-invariant subspace of V, and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ where $\mathcal{B}_i$ is a basis for V_i.

Problem 157 What property would a basis $\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ for V (of dimension k) have to have so that $[L]_B$ is upper triangular, where $L : V \to V$?

5.2 Inner product spaces

For an arbitrary vector space V over the field $\mathbb{F}$, $\langle \cdot | \cdot \rangle : V \times V \to \mathbb{F}$ is an *inner product* iff

(i) $\langle \cdot | \cdot \rangle$ is linear in the left coordinate, i.e.,

$$\langle \alpha \boldsymbol{v} + \beta \boldsymbol{u} | \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{v} | \boldsymbol{w} \rangle + \beta \langle \boldsymbol{u} | \boldsymbol{w} \rangle.$$

(ii) $\langle \cdot | \cdot \rangle$ is conjugate symmetric, that is,

$$\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \langle \boldsymbol{v} | \boldsymbol{u} \rangle^*.$$

(iii) $\langle\cdot|\cdot\rangle$ is positive,

$$\langle \boldsymbol{u}|\boldsymbol{u}\rangle \in \mathbb{R}^+ = [0,\infty) \quad \langle \boldsymbol{u}|\boldsymbol{u}\rangle = 0 \Leftrightarrow \boldsymbol{u} = \boldsymbol{0}$$

It follows that $\langle\cdot|\cdot\rangle$ is right conjugate linear:

$$\langle \boldsymbol{u}|\alpha\boldsymbol{v}+\beta\boldsymbol{w}\rangle = \alpha^*\langle \boldsymbol{u}|\boldsymbol{v}\rangle + \beta^*\langle \boldsymbol{u}|\boldsymbol{w}\rangle$$

This is a simple calculation:

$$\begin{aligned}\langle \boldsymbol{u}|\alpha\boldsymbol{v}+\beta\boldsymbol{w}\rangle &= \langle \alpha\boldsymbol{v}+\beta\boldsymbol{w}|\boldsymbol{u}\rangle^* \\ &= (\alpha\langle \boldsymbol{v}|\boldsymbol{u}\rangle + \beta\langle \boldsymbol{w}|\boldsymbol{u}\rangle)^* \\ &= \alpha^*\langle \boldsymbol{v}|\boldsymbol{u}\rangle^* + \beta^*\langle \boldsymbol{w}|\boldsymbol{u}\rangle^* \\ &= \alpha^*\langle \boldsymbol{u}|\boldsymbol{v}\rangle + \beta^*\langle \boldsymbol{u}|\boldsymbol{w}\rangle\end{aligned}$$

From here on the notation $\langle\cdot|\cdot\rangle$ will be interpreted by context, by default it means the standard inner product $\langle \boldsymbol{u}|\boldsymbol{v}\rangle = \boldsymbol{v}^*\boldsymbol{u}$ unless the context suggests otherwise. If two inner products are required at the same time, then additional notation will be used like $\langle\cdot|\cdot\rangle_1$ and $\langle\cdot|\cdot\rangle_2$.

The positive definiteness of an inner product $\langle\cdot|\cdot\rangle$ allows *length* or *magnitude* to be defined with respect to this inner product, namely,

$$\|\boldsymbol{u}\| = \sqrt{\langle \boldsymbol{u}|\boldsymbol{u}\rangle}.$$

One of the most important facts about an inner product is the *Cauchy-Schwartz* inequality:

$$|\langle \boldsymbol{u}|\boldsymbol{v}\rangle| \le \|\boldsymbol{u}\|\|\boldsymbol{v}\|$$

Problem 158 Prove the Cauchy-Schwartz inequality by first verifying it for unit vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$ and then generalizing.

Hint: Compute $\langle \boldsymbol{u}-\lambda\boldsymbol{v}|\boldsymbol{u}-\lambda\boldsymbol{v}\rangle$ for an arbitrary λ and then choose λ wisely. (See Problem 22).

Cauchy-Schwartz implies the *triangle inequality*

$$\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$$

Problem 159 Prove the triangle inequality.

The *distance* between two points in an inner product space $(V, \langle\cdot|\cdot\rangle)$ is defined as $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|$.

Problem 160 Verify that $\operatorname{dist} : V \times V \to [0, \infty)$ satisfies

(i) $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) \geq 0$ and $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) = 0$ iff $\boldsymbol{u} = \boldsymbol{v}$.

(ii) $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{dist}(\boldsymbol{v}, \boldsymbol{u})$.

(iii) (Triangle inequality) $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) \leq \operatorname{dist}(\boldsymbol{u}, \boldsymbol{w}) + \operatorname{dist}(\boldsymbol{w}, \boldsymbol{v})$.

Cauchy-Schwarz also allows the following definition of the angle between two vectors:

$$\theta_{\boldsymbol{u},\boldsymbol{v}} \stackrel{\text{df}}{=} \arccos\left(\frac{\langle \boldsymbol{u}|\boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}\right)$$

The magnitude and angle between vectors determined by an arbitrary inner product will differ from the standard inner product and this leads to different "geometries".

As expected, if $(V, \langle\cdot|\cdot\rangle)$ is an arbitrary inner product space, $\boldsymbol{u}$ and $\boldsymbol{v}$ are called *orthogonal* if $\langle\boldsymbol{u}|\boldsymbol{v}\rangle = 0$.

Problem 161 Verify the Pythagorean Theorem in an arbitrary inner product space $(V, \langle\cdot|\cdot\rangle)$. Show that
$\langle\boldsymbol{u}|\boldsymbol{v}\rangle = 0 \Rightarrow \|u+v\|^2 = \|u\|^2 + \|v\|^2$.

For V finite dimensional the Gramm-Schmidt procedure provides a method for converting any basis into an orthogonal or orthonormal basis. So if W is a subspace of V with basis $\mathcal{B} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_s\}$, then we may convert $\mathcal{B}$ into an orthonormal basis $\mathcal{B}' = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s\}$ and given $\boldsymbol{w} \in W$ we have

$$\boldsymbol{w} = \sum_{i=1}^{s} \langle\boldsymbol{w}|\boldsymbol{q}_i\rangle\boldsymbol{q}_i,$$

that is, $\boldsymbol{w}_{\mathcal{B}'} = \begin{bmatrix} \langle\boldsymbol{w}|\boldsymbol{q}_1\rangle \\ \langle\boldsymbol{w}|\boldsymbol{q}_2\rangle \\ \vdots \\ \langle\boldsymbol{w}|\boldsymbol{q}_s\rangle \end{bmatrix}$.

This is because if $\boldsymbol{w} = \sum_{i=1}^{s} \alpha_i \boldsymbol{q}_i$, then
$\langle\sum_{i=1}^{s} \alpha_i \boldsymbol{q}|\boldsymbol{q}_j\rangle = \sum_{i=1}^{s} \alpha_i \langle\boldsymbol{q}_i|\boldsymbol{q}_j\rangle = \alpha_j$ so $\alpha_j = \langle\boldsymbol{w}|\boldsymbol{q}_j\rangle$

Similarly, given any $\boldsymbol{v} \in V$, define

$$\begin{aligned} \operatorname{proj}_W(\boldsymbol{v}) &\stackrel{\text{df}}{=} \sum_{i=1}^{s} \langle\boldsymbol{v}|\boldsymbol{q}_i\rangle\boldsymbol{q}_i \\ \operatorname{proj}_W^{\perp}(\boldsymbol{v}) &\stackrel{\text{df}}{=} \boldsymbol{v} - \operatorname{proj}_W(\boldsymbol{v}) \end{aligned}$$

As before $\operatorname{proj}_W^{\perp}(\boldsymbol{v}) \perp W$, for let $\boldsymbol{w} = \sum_{i=1}^{s} \alpha_i \boldsymbol{q}_i \in W$, then

$$\begin{aligned}
\left\langle \boldsymbol{v} - \sum_{i=1}^{s} \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle \boldsymbol{q}_i \middle| \sum_{j=1}^{s} \alpha_j \boldsymbol{q}_j \right\rangle &= \left\langle \boldsymbol{v} \middle| \sum_{i=1}^{s} \alpha_i \boldsymbol{q}_i \right\rangle - \sum_{i=1}^{s} \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle \left\langle \boldsymbol{q}_i \middle| \sum_{j=1}^{s} \alpha_j \boldsymbol{q}_j \right\rangle \\
&= \sum_{i=1}^{s} \alpha_i^* \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle - \sum_{i=1}^{s} \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle \sum_{j=1}^{s} \alpha_j^* \langle \boldsymbol{q}_i | \boldsymbol{q}_j \rangle \\
&= \sum_{i=1}^{s} \alpha_i^* \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle - \sum_{i=1}^{s} \langle \boldsymbol{v} | \boldsymbol{q}_i \rangle \alpha_i^* = 0
\end{aligned}$$

We may now use Pythagorean Theorem as before to see that $\boldsymbol{p} = \operatorname{proj}_W(\boldsymbol{v})$ is the unique element of W that minimizes $\operatorname{dist}(\boldsymbol{p}, \boldsymbol{v})$. For let $\boldsymbol{w} \in W$ be distinct from $\boldsymbol{p}$, then

$$\begin{aligned}
\operatorname{dist}(\boldsymbol{w}, \boldsymbol{v})^2 &= \|\boldsymbol{w} - \boldsymbol{v}\|^2 \\
&= \|(\boldsymbol{w} - \boldsymbol{p}) + (\boldsymbol{p} - \boldsymbol{v})\|^2 \\
&= \|\boldsymbol{w} - \boldsymbol{p}\|^2 + \|\boldsymbol{p} - \boldsymbol{v}\|^2 \\
&> \|\boldsymbol{p} - \boldsymbol{v}\|^2 = \operatorname{dist}(\boldsymbol{p}, \boldsymbol{v})^2
\end{aligned}$$

This shows $\boldsymbol{p}$ is the **unique point** in W minimizing $\operatorname{dist}(\boldsymbol{p}, \boldsymbol{v})$.

Problem 162 Show that $V = W \oplus W^{\perp}$ for any finite dimensional inner product space $(V, \langle \cdot | \cdot \rangle)$.

Problem 163 Let $L \in \mathcal{L}(V, W)$ with V finite dimensional. Show that $\dim(\operatorname{Ker}(L)^{\perp}) = \dim(\operatorname{Img}(L))$. Again this is analogous to $\dim(\operatorname{RS}(A)) = \dim(\operatorname{CS}(A))$.)

The next exercise gives what a large number of examples of "non-standard" inner products on $\mathbb{C}^n$. Later, see Problem 167, it is shown that all non-standard inner products are generated this way.

Problem 164 Given any positive Hermitian matrix A, i.e., $A^* = A$ and $\boldsymbol{x}^* A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$, $\boldsymbol{y}^* A \boldsymbol{x} = \langle \boldsymbol{x} | \boldsymbol{y} \rangle_A$ is an inner product on $\mathbb{C}^n$.

For example if $w_i \in \mathbb{R}$ and $w_i > 0$ for $i = 1, \dots, n$, then $\langle \boldsymbol{u}|\boldsymbol{v}\rangle_{\boldsymbol{w}} = \sum_{i=1}^{n} w_i v_i^* u_i$ is the *weighted inner product with weight* $\boldsymbol{w}$. This is an example of the preceding by taking $A = \mathrm{diag}(w_1, \dots, w_n)$, the $n \times n$ matrix with
$$A_{ij} = \begin{cases} w_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Problem 165 Consider the weighted inner product on $\mathbb{R}^2$ given by $\langle \boldsymbol{u}|\boldsymbol{v}\rangle_{\boldsymbol{w}} = w_1 v_1 u_1 + w_2 v_2 u_2$. Describe geometrically the unit circle under this inner product, that is, what shape is $\{\boldsymbol{u} \mid \|\boldsymbol{u}\| = 1\}$ where $\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}|\boldsymbol{u}\rangle_{\boldsymbol{w}}$.

Problem 166 Let $l : \mathbb{F}^n \to \mathbb{F}$ be linear, show that there is a unique $\boldsymbol{z}_l \in \mathbb{F}^n$ so that $l(\boldsymbol{x}) = \boldsymbol{z}_l{}^* \boldsymbol{x}$.

Problem 167 Let $\langle\cdot|\cdot\rangle$ be any inner product on $\mathbb{F}^n$, the interesting case being something other than the standard inner product, then there is a positive Hermitian matrix A such that $\langle \boldsymbol{x}|\boldsymbol{y}\rangle = \boldsymbol{y}^* A\boldsymbol{x}$.

Problem 168 Show that for every inner product $\langle\cdot|\cdot\rangle$ on $\mathbb{F}^n$ there is a one-to-one and onto linear $K : \mathbb{F}^n \to \mathbb{F}^n$ so that
$\langle \boldsymbol{u}|\boldsymbol{v}\rangle = (K\boldsymbol{v})^*(K\boldsymbol{u})$. So K transforms the standard inner product into the given inner product.

The preceding problems demonstrate the possibilities of inner products on $\mathbb{F}^n$, the following now demonstrates the generality of the concept.

The "next" vector spaces to consider after $\mathbb{F}^n$ are $M_{mn}(\mathbb{F})$, viewing each $m \times n$ matrix as a vector in $\mathbb{F}^{mn}$ we get the induced inner product and corresponding norm, called the *Frobenious norm*, the following characterizes this in a slightly different fashion.

Problem 169 Define $\langle A|B\rangle = \operatorname{tr}(B^*A)$ on $M_{mn}(\mathbb{F})$.

(a) Show this is an inner product.

(b) Compute $\|A\|$ for $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$.

(c) Compute $\cos(\theta_{A,B})$ for A as in (b) and $B = \begin{bmatrix} -1 & 2 & 3 \\ 0 & 2 & 0 \end{bmatrix}$.

(d) Find A', B' so that $\mathrm{Span}(A, B) = \mathrm{Span}(A, B')$ and the set $\{A', B'\}$ is orthonormal.

(e) Show that the norm induced by this inner product is the Frobenius norm.

Problem 170 Let $V = C([0, 1], \mathbb{R})$ be the vector space of all continuous functions $f : [0, 1] \to \mathbb{R}$. Define

$$\langle f|g\rangle = \int_0^1 fg\,dx$$

(a) Show that this is an inner product.

(b) Determine $\|2 + x - x^2\|$.

(c) Determine $\langle 2 + x - x^2 | 1 + x \rangle$.

(d) Determine $\cos(\theta)$ where θ is the angle between $2 + x - x^2$ and $1 + x$.

(e) Determine $\operatorname{dist}(2 + x - x^2, 1 + x)$.

(f) Find an orthonormal basis for $W = \text{Span}(2 + x - x^2, 1 + x)$.

(g) Find the orthogonal projection of $1 - x$ in W.

Problem 171 Consider $C([-\pi, \pi], \mathbb{R})$. Show that $f_n = \frac{1}{\sqrt{\pi}} \cos(nx)$, for $n \geq 0$ an integer, and $g_m = \frac{1}{\sqrt{\pi}} \sin(mx)$ for m an integer > 0 are all mutually orthonormal. That is $\langle f_n | g_m \rangle = 0$ for all n and $m \neq 0$, $\langle f_n | f_m \rangle = 0$ for all $m \neq n$, $\langle f_n | f_m \rangle = 0$ for all $m \neq n$ with $m \neq 0 \neq n$, and $\langle f_n | f_n \rangle = \langle g_n | g_n \rangle = 1$. [17]

Let

$$a_n(f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)\, dx \text{ for } n \geq 0 \text{ and}$$
$$b_n(f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)\, dx \text{ for } n > 0,$$

then for $f \in C([-\pi, \pi], \mathbb{R})$ (this is non-trivial)

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \langle f | f_n \rangle f_n(x)\, dx + \sum_{n=1}^{\infty} \langle f | g_n \rangle g_n(x)\, dx \\ &= \frac{1}{2} a_0(f) + \sum_{n=1}^{\infty} a_n(f) \cos(nx) + \sum_{n=1}^{\infty} b_n(x) \sin(nx) \end{aligned}$$

[17] A generalization of this would be to consider $C([-\pi, \pi], \mathbb{C})$ and take $f_n(x) = e^{inx} = \cos(nx) + i \sin(nx)$ for all integers n.

5.2.1 Dual space and adjoints

The adjoint of a linear function can be generalized given that for any subspace W of an inner product space V, V can be decomposed as $V = W \oplus W^{\perp}$, see Problem 162 for finite dimensional V. For any linear $L : V \to W$, where V and W are inner product spaces with inner products $\langle\cdot|\cdot\rangle_V$ and $\langle\cdot|\cdot\rangle_W$, there is a unique linear $L^* : W \to V$ satisfying

$$\langle L(\boldsymbol{v})|\boldsymbol{w}\rangle_W = \langle\boldsymbol{v}|L^*(\boldsymbol{w})\rangle_V.$$

This leads a generalization of the orthogonal projection operator and the consequent least squares solutions to $L(\boldsymbol{v}) = \boldsymbol{b}$, namely the orthogonal projection map is $P = L(L^*L)^{-1}L^*$. Other consequences of the existence of an adjoint follow as well.

Finally the picture of the four fundamental subspaces associated to a linear operator emerges exactly as we had for matrices.

We can generalize Problem 166. One generalization would be:

Let $\langle\cdot|\cdot\rangle$ be an arbitrary inner product on $\mathbb{F}^n$ and let $l : \mathbb{F}^n \to \mathbb{F}$ be linear. Then there is unique $\boldsymbol{z}_l \in \mathbb{F}^n$ so that $l(\boldsymbol{x}) = \langle\boldsymbol{x}|\boldsymbol{z}_l\rangle$.

More generally we have:

The *dual space* to a vector space V is the space $\mathcal{L}(V, \mathbb{F})$ of linear functions into $\mathbb{F}$. This space is denoted V^* and typically objects in V^* are denoted $\boldsymbol{v}^*$.

Problem 172 Show that if $(V, \langle\cdot|\cdot\rangle)$ is an inner product space, then for any $\boldsymbol{v}^* \in V^*$, there is a unique $\boldsymbol{y}_{\boldsymbol{v}^*}$ so that for all $\boldsymbol{v} \in V$,

$\boldsymbol{v}^*(\boldsymbol{v}) = \langle \boldsymbol{v} | \boldsymbol{y}_{\boldsymbol{v}^*} \rangle$. Moreover show that, the map $\boldsymbol{v}^* \mapsto \boldsymbol{y}_{\boldsymbol{v}^*}$ is a conjugate linear one-to-one and onto map.

Hint: If $\boldsymbol{v}^* = \boldsymbol{0}_{V^*}$, then take $\boldsymbol{y}_{V^*} = \boldsymbol{0}_V$. Else, take $\boldsymbol{w} \in \operatorname{Ker}(\boldsymbol{v}^*)^{\perp}$[18], notice that for all $\boldsymbol{u} \in V$:

$$\boldsymbol{v}^*(\boldsymbol{u})\boldsymbol{w} - \boldsymbol{v}^*(\boldsymbol{w})\boldsymbol{u} \in \operatorname{Ker}(\boldsymbol{v}^*)$$

So

$$\langle \boldsymbol{v}^*(\boldsymbol{u})\boldsymbol{w} - \boldsymbol{v}^*(\boldsymbol{w})\boldsymbol{u} | \boldsymbol{w} \rangle = \boldsymbol{v}^*(\boldsymbol{u})\langle \boldsymbol{w} | \boldsymbol{w} \rangle - \boldsymbol{v}^*(\boldsymbol{w})\langle \boldsymbol{u} | \boldsymbol{w} \rangle = 0$$

The following is the generalization of adjoint to arbitrary inner product spaces:

Problem 173 Show that for $(V, \langle \cdot | \cdot \rangle_V)$ and $(W, \langle \cdot | \cdot \rangle_W)$ be inner product spaces and $L \in \mathcal{L}(V, W)$. There is a unique $L^* \in \mathcal{L}(W, V)$ so that $\langle L(\boldsymbol{v}) | \boldsymbol{w} \rangle_W = \langle \boldsymbol{v} | L^*(\boldsymbol{w}) \rangle_W$.

[18] What follows depends on $V = \operatorname{Ker}(\boldsymbol{v}^*)^{\perp} \oplus \operatorname{Ker}(\boldsymbol{v}^*)$ and that $\dim(\operatorname{Ker}(\boldsymbol{v}^*)^{\perp}) = \dim(\operatorname{Img}(\boldsymbol{v}^*)) = 1$. These follow from Problem 162 under the additional assumption that V is finite dimensional. In the general case one needs some additional assumptions to know that $V = \operatorname{Ker}(L)^{\perp} \oplus \operatorname{Ker}(L)$.

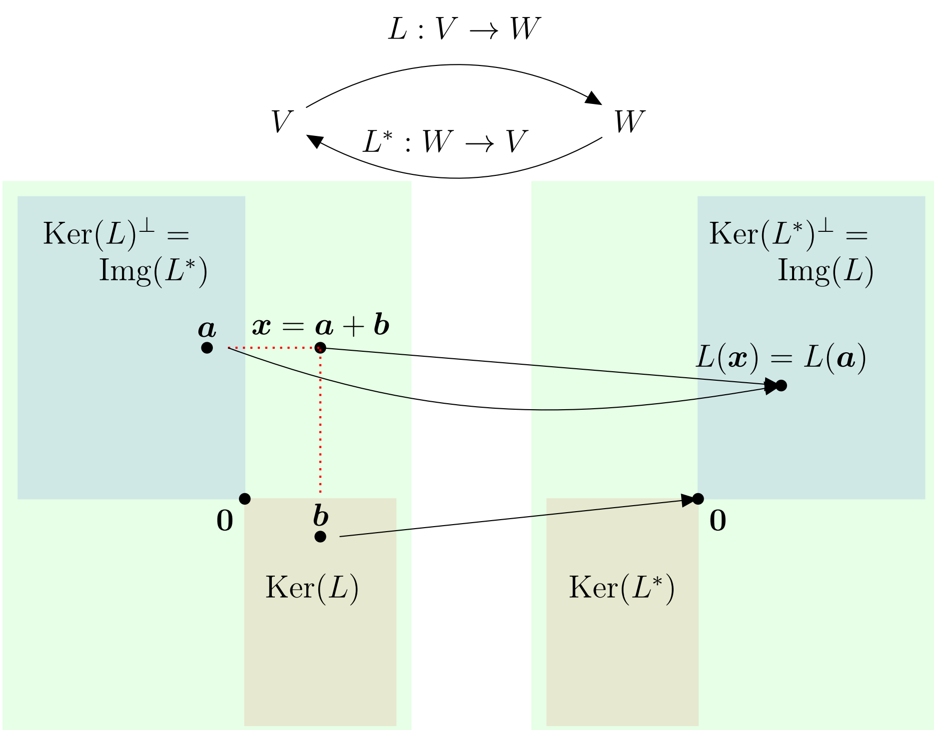
Problem 174 Let V, W, L, and L^* be as in Problem 173. Show that $\mathrm{Img}(L^*) = \mathrm{Ker}(L)^{\perp}$. (This is like $\mathrm{CS}(A^*) = \mathrm{NS}(A)^{\perp}$, notice that for $A \in M_{mn}(\mathbb{C})$ this is the correct equation since $\mathrm{row}_i(A)\boldsymbol{x} = \mathrm{col}_i(A^*)^*\boldsymbol{x} = \langle \boldsymbol{x} | \mathrm{col}_i(A^*) \rangle$. For A with complex entries, $\mathrm{RS}(A) \neq \mathrm{CS}(A^*)$, rather $\mathrm{RS}(A)^* = \mathrm{CS}(A^*)$ under the obvious definition.)

Problem 175 Let V, W, L, and L^* be as in Problem 173. Show that $L^{**} = L$, $(LS)^* = S^*L^*$, and $(L^*)^{-1} = (L^{-1})^*$ when L is invertible.

Problem 176 Let V, W, L, and L^* be as in Problem 173. Show that if L is one-to-one, then L^*L is invertible.

Problem 177 Let V, W, L, and L^* be as in Problem 173. Show that if L is one-to-one, then $L(L^*L)^{-1}L^* = P$ is the orthogonal projection function onto $\mathrm{Img}(L)$.

Finally, the picture of the four fundamental subspaces associated to a linear function



6 Eigenvalues, eigenvectors, and diagonalization

In this section we decompose the action of a matrix into scalar multiplication along certain *principal axes* or *eigenspaces*. Given a linear mapping $L : V \to V$ a scalar λ is an *eigenvalue* for L iff $L(\boldsymbol{v}) = \lambda\boldsymbol{v}$ for some $\boldsymbol{v} \neq \mathbf{0}_V$. If λ is an eigenvalue for L, and $L(\boldsymbol{v}) = \lambda\boldsymbol{v}$, then $\boldsymbol{v}$ is an associated *eigenvector*.

If L is given by a matrix A and $V = \mathbb{R}^n$, then say λ is an eigenvalue for A iff $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some non-zero $\boldsymbol{x} \in \mathbb{R}^n$ and call such an $\boldsymbol{x}$ an eigenvector associated to λ.

Problem 178 Let $E_\lambda = \{\boldsymbol{v} \mid L(\boldsymbol{v}) = \lambda\boldsymbol{v}\}$ for λ and eigenvalue of L. Show that E_λ is an invariant subspace of V.

Problem 179 For A an $n \times n$ matrix, show that $E_0 = \mathrm{NS}(A)$ and consequently, A is singular iff 0 is an eigenvalue of A.

Problem 180 Suppose P is a projection matrix what are the eigenvalues/eigenspaces of P? Suppose P is an orthogonal projection matrix what else can you say about the eigenspaces of P? (We will show below that the eigenspaces for a symmetric/hermitian matrix are orthogonal.)

Problem 181 Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be reflection about the line $y = x$. L is given by the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ what are the eigenvalues/eigenspaces?

Problem 182 Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation counterclockwise through the angle θ. For what θ are there real eigenvalues? When there are real eigenvalues what are the eigenspaces?

Problem 183 Show that if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues for L and $\boldsymbol{v}_i \in E_{\lambda_i}$, then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ are linearly independent.

The previous problem shows that if $\lambda_1, \ldots, \lambda_s, \lambda_{s+1}, \ldots, \lambda_k$ are distinct eigenvalues, then

$$\left(E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_s}\right) \cap \left(E_{\lambda_{s+1}} \oplus \cdots \oplus E_{\lambda_k}\right) = \{\mathbf{0}\}.$$

The next exercise follows easily.

Problem 184 Show that if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues for L and $\mathcal{B}_i$ is a basis for E_{λ_i}, then $\mathcal{B} = \bigcup_{i=1}^{k} \mathcal{B}_i$ is linearly independent.

Define the *geometric multiplicity of the eigenvalue* λ to be $\dim(E_\lambda)$. If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of $L : V \to V$ where $\dim(V) = n$, then

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k}$$

has dimension $\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k}$. The linear operation L, or matrix A, is said to be *diagonalizable* if the sum of the geometric multiplicities of the eigenvalues is equal to $\dim(V)$. This is equivalent to V having a basis of eigenvectors.

If $\boldsymbol{v}_1, \cdots \boldsymbol{v}_n$ is a basis of eigenvectors for $L : \mathbb{R}^n \to \mathbb{R}^n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ let

$$S = \left[\boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right] \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

S is called the eigenvector matrix and the diagonal matrix Λ is called the eigenvalue matrix. We clearly have

$$\begin{aligned} AS &= \left[A\boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, A\boldsymbol{v}_n\right] = \left[\lambda_1\boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, \lambda_n\boldsymbol{v}_n\right] \\ &= \left[\boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = S\Lambda \end{aligned}$$

Since the columns of S are independent S^{-1} exists and

$$A = S\Lambda S^{-1}$$

This decomposition is called a *diagonalization* of A.

Problem 185 Show that is $\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ is a basis of eigenvalues with $L(\boldsymbol{u}_i) = \lambda_i \boldsymbol{u}_i$, then

$$[L]_{\mathcal{B}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

so $[L] = B[L]_{\mathcal{B}} B^{-1}$ is viewed in terms of a change of basis.

Problem 186 Show that if $A = S\Lambda S^{-1}$ is a diagonalizaion of A, then S is a matrix of eigenvectors for A and Λ is the associated matrix of eigenvalues.

Problem 187 Given that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is a basis of eigenvectors for L with associated eigenvalues $\lambda_1, \ldots, \lambda_n$, express $L^k(\boldsymbol{u})$ given that $\boldsymbol{u} = \sum_{i=1}^n \alpha_i \boldsymbol{v}_i$.

There are many useful properties of a diagonalization of A which we will investigate, one of the most useful is the ease with which we can take powers a diagonalized matrix.

If $A = S\Lambda S^{-1}$, then $A^k = S\Lambda^k S^{-1}$, where Λ^k is easy to compute since $(\Lambda^k)_{ij} = (\Lambda_{ij})^k$.

Problem 188 Show that if λ is an eigenvalue of L and $\boldsymbol{v}$ an

associated eigenvector, then λ^k is an eigenvalue for L^k and $\boldsymbol{v}$ is an eigenvector for L^k associated to λ^k.

Problem 189 This is important in studying dynamical systems for example if initially 30% of people use brand X, 25% brand Y, and 45% brand Z is expressed by $\boldsymbol{u}_0 = \begin{bmatrix} .3 \\ .25 \\ .45 \end{bmatrix}$ and at time $t+1$ we have

$$\boldsymbol{u}_{t+1} = \begin{bmatrix} .2 & .3 & .5 \\ .6 & .1 & .4 \\ .2 & .6 & .1 \end{bmatrix} \boldsymbol{u}_t$$

then 20% of brand X stay with brand X, 60% of brand X users switch to brand Y, and 20% of brand X users switch to brand Z, etc. The question is: "What is the eventual distribution of users?" That is what is $\lim_{t\to\infty} \boldsymbol{u}_t$?

If there is such a limit, then it should be an eigenvector for 1 and its entries should sum to 1. show that such a vector exists and find it. Later we will diagonalize the matrix and find its limit.

Problem 190 Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation so that

$$L\left(\begin{bmatrix}1\\2\\-1\end{bmatrix}\right) = \begin{bmatrix}3\\6\\-3\end{bmatrix} \quad L\left(\begin{bmatrix}-2\\4\\-2\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\1\end{bmatrix} \quad L\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}0\\0\\-5\end{bmatrix}$$

compute $L^5\left(\begin{bmatrix}-1\\6\\-2\end{bmatrix}\right)$. Find $[L]$ and $[L^5]$.

Problem 191 Show that every $L \in \mathcal{L}(V)$, for V a finite dimensional vector space over $\mathbb{C}$ has an eigenvalue.

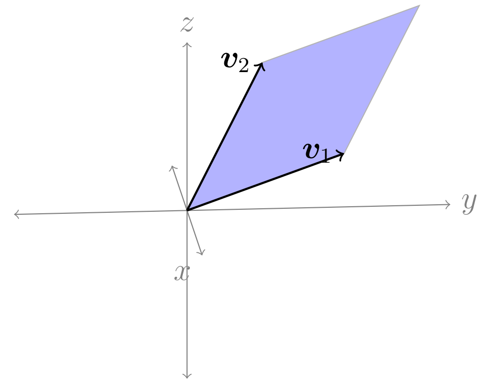


Figure 1: A 2-parallelepiped in $\mathbb{R}^3$.

6.1 Determinants and volume

Under Revision!!!

A *parallelogram* (or 2-parallelepiped) in $\mathbb{R}^n$ is determined by 2 non-co-linear vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ with the 4 vertices being $\mathbf{0}$, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_1 + \boldsymbol{v}_2$:

a parallelepiped (or 3-parallelepiped) in $\mathbb{R}^n$ is determined by 3 non-coplanar vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ and has vertices $\mathbf{0}$, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_1 + \boldsymbol{v}_2$, $\boldsymbol{v}_1 + \boldsymbol{v}_2$, $\boldsymbol{v}_2 + \boldsymbol{v}_3$, and $\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3$.

An m-parallelepiped in n-space is determined by m linearly independent vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ in $\mathbb{R}^n$. For each subset S of $\{1, \ldots, m\}$ there is a vertex P_S at $P + \sum_{i \in S} \boldsymbol{u}_i$. The

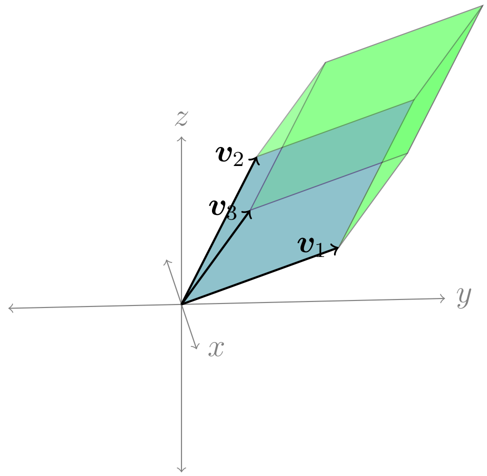


Figure 2: A 3-parallelepiped in $\mathbb{R}^3$

m-parallelepiped determined by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ will be denoted $P(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$. Formally

$$P(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m) \stackrel{\text{df}}{=} \left\{ \sum_{i=1}^{m} a_1 \boldsymbol{u}_i \,\middle|\, 0 \leq a_i \leq 1, \sum_{i=1}^{n} a_i \leq 1 \right\}.$$

Define addition of m-parallelepipeds as follows:

$$\begin{aligned} P(\boldsymbol{u}_1, \ldots, &\boldsymbol{u}_{i-1}, \boldsymbol{u}_i, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_m) + P(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_i', \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_m) \\ &\stackrel{\text{df}}{=} P(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_i + \boldsymbol{u}_i', \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_m) \end{aligned}$$

Define a function $\Delta : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ so that the volume of the n-parallelepiped $P(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ is $|\Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)|$. This makes sen

n-parallelepiped ⧉$(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ will be denoted $\Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$. The volume should be the absolute value of the signed volume:

$$\left| \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) \right| = \operatorname{Vol}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n).$$

where $\operatorname{Vol}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ is the n-dimensional volume.

The following three properties of $\Delta : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ completely determine its value.

(1) (*multilearity*)

$$\begin{gathered} \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \alpha \boldsymbol{u}_i + \beta \boldsymbol{u}_i', \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n) \\ \alpha \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_i, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n) \\ + \beta \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_i', \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n), \end{gathered}$$

for each i.

(2) If the $\boldsymbol{u}_i$'s are not linearly independent, then $\Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) = 0$ since the parallelepiped would not be n-dimensional. This is captured simply by

$$\Delta(\ldots, \boldsymbol{u} \ldots, \boldsymbol{u}, \ldots) = 0.$$

(3) It is clear that the volume of the unit cube should be 1, that is,

$$\operatorname{Vol}(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = 1.$$

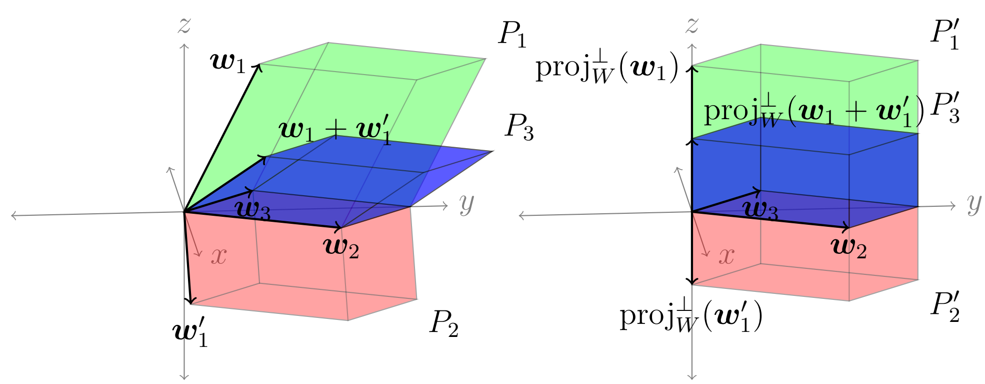
Hence, either $\Delta(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = 1$ or $\Delta(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = -1$, by fiat choose the former. This single choice "orients" the n-dimensional space in a certain fashion.

$$\Delta(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = 1 \tag{2}$$

To see why Δ is multilinear let W be the $(n-1)$-dimensional hyperplane determined by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n$. Let $\boldsymbol{n}$ be normal to the plane and unit length and let $\operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u}) = \boldsymbol{u} \cdot \boldsymbol{n}$ so that $\operatorname{proj}_W^{\perp}(\boldsymbol{u}) = \operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u})\boldsymbol{n}$.

$$\begin{aligned}
&\Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_i, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n) \\
&\qquad = \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \operatorname{proj}_W^{\perp}(\boldsymbol{u}_i), \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n) \\
&\qquad = \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u}_i)\boldsymbol{n}, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n) \\
&\qquad = \operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u}_i)\Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{n}, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n)
\end{aligned}$$

This together with the fact that $\operatorname{comp}_{\boldsymbol{n}} : \mathbb{R}^n \to \mathbb{R}$ is linear, i.e., $\operatorname{comp}_{\boldsymbol{n}}(\alpha\boldsymbol{u}_i + \beta\boldsymbol{u}_i') = \alpha\operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u}_i) + \beta\operatorname{comp}_{\boldsymbol{n}}(\boldsymbol{u}_i')$ gives the desired multilinearity of Δ.



Illustrated is the case of adding two parallelepipes to form a third

$$\begin{aligned}
P_1 &= \square(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \\
P_2 &= \square(\boldsymbol{w}_1', \boldsymbol{w}_2, \boldsymbol{w}_3) \\
P_3 &= \square(\boldsymbol{w}_1 + \boldsymbol{w}_1', \boldsymbol{w}_2, \boldsymbol{w}_3) \\
&= \square(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) + \square(\boldsymbol{w}_1', \boldsymbol{w}_2, \boldsymbol{w}_3)
\end{aligned}$$

The sign of P_1 and P_3 are the same and opposite that of P_2 due to being on different sides of the plane $W = \operatorname{Span}\{\boldsymbol{w}_2, \boldsymbol{w}_3\}$.

From $\Delta(\ldots, \boldsymbol{u}, \ldots, \boldsymbol{u}, \ldots) = 0$ we get

$$\Delta(\ldots, \boldsymbol{u}, \ldots, \boldsymbol{v}, \ldots) = -\Delta(\ldots, \boldsymbol{v}, \ldots, \boldsymbol{u}, \ldots),$$

that is, swapping, two vectors changes the sign. The proof is a simple calculation:

$$\Delta(\ldots,\boldsymbol{u},\ldots,\boldsymbol{v},\ldots)+\Delta(\ldots,\boldsymbol{v},\ldots,\boldsymbol{u},\ldots)=0.$$

Just calculate:

$$\begin{aligned}\Delta(\ldots,&\boldsymbol{u},\ldots,\boldsymbol{v},\ldots)+\Delta(\ldots,\boldsymbol{v},\ldots,\boldsymbol{u},\ldots)\\ &=\Delta(\ldots,\boldsymbol{u},\ldots,\boldsymbol{v},\ldots)+\Delta(\ldots,\boldsymbol{u},\ldots,\boldsymbol{u},\ldots)+\\ &\quad\Delta(\ldots,\boldsymbol{v},\ldots,\boldsymbol{u},\ldots)+\Delta(\ldots,\boldsymbol{v},\ldots,\boldsymbol{v},\ldots)\\ &=\Delta(\ldots,\boldsymbol{u},\ldots,\boldsymbol{v}+\boldsymbol{u},\ldots)+\Delta(\ldots,\boldsymbol{v},\ldots,\boldsymbol{v}+\boldsymbol{u},\ldots)\\ &=\Delta(\ldots,\boldsymbol{u}+\boldsymbol{v},\ldots,\boldsymbol{v}+\boldsymbol{u},\ldots)=0.\end{aligned}$$

Each step just follows by multilinearity. Similarly,

$$\Delta(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)=0\Leftrightarrow\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n \text{ are not linearly independent}^{19}.$$

Claim The properties (1) and (2) completely decide the value of $\Delta(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$. $\diamondsuit$

First consider two vectors in $\mathbb{R}^2$:

In $\mathbb{R}^2$ let $v_1=\langle v_{11},v_{12}\rangle=v_{11}\boldsymbol{e}_1+v_{12}\boldsymbol{e}_2$, similarly for $\boldsymbol{v}_2$.

$$\begin{aligned}\Delta(\boldsymbol{v}_1,\boldsymbol{v}_2)&=\Delta(v_{11}\boldsymbol{e}_1+v_{12}\boldsymbol{e}_2,\boldsymbol{v}_2)\\ &=v_{11}\Delta(\boldsymbol{e}_1,\boldsymbol{v}_2)+v_{12}\Delta(\boldsymbol{e}_2,\boldsymbol{v}_2)\\ &=v_{11}\Delta(\boldsymbol{e}_1,v_{21}\boldsymbol{e}_1+v_{22}\boldsymbol{e}_2)+v_{12}\Delta(\boldsymbol{e}_2,v_{21}\boldsymbol{e}_1+v_{22}\boldsymbol{e}_2)\\ &=v_{11}v_{21}\Delta(\boldsymbol{e}_1,\boldsymbol{e}_1)+v_{1,1}v_{22}\Delta(\boldsymbol{e}_1,\boldsymbol{e}_2)+v_{12}v_{21}\Delta(\boldsymbol{e}_2,\boldsymbol{e}_1)+v_{12}v_{22}\Delta(\boldsymbol{e}_2,\boldsymbol{e}_2)\\ &=v_{11}v_{21}(0)+v_{11}v_{22}(1)+v_{12}v_{21}(-1)+v_{12}v_{22}(0)\\ &=v_{11}v_{22}-v_{12}v_{21}\end{aligned}$$

Problem 192 Show that $\Delta(\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3)=u_{11}\Delta((u_{22},u_{23}),(u_{32},u_{33}))-u_{12}\Delta((u_{21},u_{23}),(u_{31},u_{33}))+u_{13}\Delta((u_{21},u_{22}),(u_{31},u_{32}))$. This gives the usual expansion of a 3×3 determinant.

In general, in order to get the usual expansion of determinants one argues as follows:

$$\Delta(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \sum_{i=1}^{n} v_{1\,i} \Delta(\boldsymbol{e}_i, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$$

Now using the above properties it follows that

$$\Delta(\boldsymbol{e}_i, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) =$$
$$(-1)^{i+1}\Delta(\boldsymbol{v}_2 - v_{2\,i}\boldsymbol{e}_i, \ldots, \boldsymbol{v}_i - v_{i\,i}\boldsymbol{e}_i, \boldsymbol{e}_i, \boldsymbol{v}_{i+1} - v_{i+1\,i}\boldsymbol{e}_i, \ldots, \boldsymbol{v}_n - v_{n\,i}\boldsymbol{e}_i)$$

Viewing the vectors as the rows of a matrix

$$\Delta \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ v_{2\,1} & v_{2\,2} & \cdots & v_{2\,i-1} & v_{2\,i} & v_{2\,i+1} & \cdots & v_{2\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ v_{i\,1} & v_{i\,2} & \cdots & v_{i\,i-1} & v_{i\,i} & v_{i\,i+1} & \cdots & v_{i\,n} \\ v_{i+1\,1} & v_{i+1\,2} & \cdots & v_{i+1\,i-1} & v_{i+1\,i} & v_{i+1\,i+1} & \cdots & v_{i+1\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ v_{n\,1} & v_{n\,2} & \cdots & v_{n\,i-1} & v_{n\,i} & v_{n\,i+1} & \cdots & v_{n\,n} \end{bmatrix} =$$

$$(-1)^{i+1}\Delta \begin{bmatrix} v_{2\,1} & v_{2\,2} & \cdots & v_{2\,i-1} & 0 & v_{2\,i+1} & \cdots & v_{2\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ v_{i\,1} & v_{i\,2} & \cdots & v_{i\,i-1} & 0 & v_{i\,i+1} & \cdots & v_{i\,n} \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ v_{i+1\,1} & v_{i+1\,2} & \cdots & v_{i+1\,i-1} & 0 & v_{i+1\,i+1} & \cdots & v_{i+1\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ v_{n\,1} & v_{n\,2} & \cdots & v_{n\,i-1} & 0 & v_{n\,i+1} & \cdots & v_{n\,n} \end{bmatrix}$$

For $\boldsymbol{u}_k$ in $\mathbb{R}^{n-1}$ let $\boldsymbol{u}'_k \in \mathbb{R}^n$ with $u'_{k\,j} = u_{k\,j}$ for $j < i$, $u'_{k\,i} = 0$, and $u'_{k,j} = u_{k,j-1}$ for $j > i$. Now define

$$\Delta'(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}) = \Delta(\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_{i-1}, \boldsymbol{e}_i, \boldsymbol{u}'_i, \ldots, \boldsymbol{u}'_n)$$

It should be clear that $\Delta' : (\mathbb{R}^{n-1})^{n-1} \to \mathbb{R}$ satisfies (1)–(3), thus $\Delta'(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}) = \Delta(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1})$ and hence,

$$\Delta(\boldsymbol{e}_i, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) = (-1)^{i+1}\Delta(\boldsymbol{v}_2^i, \ldots, \boldsymbol{v}_n^i)$$

where $\boldsymbol{v}_j^i$ is just $\boldsymbol{v}_j$ with the i^{th} component dropped. This is the usual cofactor expansion for the determinant along the first row:

$$\Delta(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \sum_{i=1}^{n} v_{1\,i}(-1)^{i+1}\Delta(\boldsymbol{v}_2^i, \ldots, \boldsymbol{v}_n^i)$$

Given an $n \times n$ matrix A, the matrix $M_{ij}(A)$ is the $(n-1) \times (n-1)$ matrix which is A with the i^{th} row and j^{th} column deleted. $M_{ij}(A)$ is called the *ij^{th} minor of* A. The ij^{th} *cofactor* of A is $\operatorname{cof}_{ij}(A) = (-1)^{(i+j)} \det M_{ij}(A)$. So the formula above for $\det(A)$ becomes:

$$\det A = \sum_{i=1}^{n} A_{1i}(-1)^{1+i} \det M_{1i}(A).$$

We can expand along the i^{th} row, instead of the first, to get

$$\det A = \sum_{k=1}^{n} A_{ik}(-1)^{i+k} \det M_{ik}(A).$$

Define the ij^{th} *cofactor* of A to be $\operatorname{cof}_{ij}(A) = (-1)^{i+j} \det M_{ij}(A)$. Thus the *cofactor* expansion of the determinant along the i^{th} row is

$$\boxed{\det(A) = \sum_{k=1}^{n} A_{ik} \operatorname{cof}_{ik}(A)}$$

Define $\operatorname{cof}(A)$ be the $n \times n$ matrix of cofactors. A direct computation shows

$$\boxed{A \operatorname{cof}(A)^T = \det(A) I}$$

and so if $\det(A) \neq 0$, i.e., A is non-singular,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{cof}(A)^T$$

Problem 193 Prove $A \operatorname{cof}(A)^T = \det(A)I$. This amounts to verifying that

$$\begin{aligned}(A \operatorname{cof}(A)^T)_{ij} &= \operatorname{row}_i(A)\operatorname{col}_j(\operatorname{cof}(A)^T) = \operatorname{row}_i(A) \cdot \operatorname{row}_j(\operatorname{cof}(A)) \\ &= \begin{cases} \det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}\end{aligned}$$

Problem 194 Show that $\det(A)$ for A triangular is $\prod_{i=1}^{n} A_{ii}$.

Problem 195 Show that $\det(AB) = \det(A)\det(B)$ for $A, B \in M_{nn}$. Split this into two cases, (1) either A is singular or (2) otherwise. If A is singular, then so is AB and hence both sides of the equality are 0. In the case that A is non-singular, A can be written as the product of elementary matrices.

The main issue is to see that $\det(EB) = \det(E)\det(B)$ for each elementary matrix E.

A similar argument shows that $\det(A) = \det(A^T)$.

Problem 196 Show that $\det(A) = \det(A^T)$. As in the previous problem use that A is singular iff A^T is singular and in this case $\det(A) = 0 = \det(A^T)$. Otherwise $A = E_1 \cdots E_k$ where the E_i are elementary matrices. Now argue $\det(E) = \det(E^T)$ for elementary matrices and use the previous problem.

This shows that we can expand determinants along rows as well as columns. Expanding on the j^{th} column gives

$$\det(A) = \sum_{k=1}^{n} A_{kj} \operatorname{cof}_{kj}(A).$$

Problem 197 Show that $\det(A)$ is the $\pm$(product of the pivots of A). Consider $A \to U$ where the only elementary row operations used are $R_i + \alpha R_j \to R_i$ and $R_i \leftrightarrow R_j$ (swap rows.)

Problem 198 For A that is non-singular argue $\det(A^{-1}) = \frac{1}{\det(A)}$.

Problem 199 Suppose $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}$ are linearly independent in $\mathbb{R}^n$ and let $\bigtimes_{i=1}^{n-1} \boldsymbol{u}_i = \det \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \\ \hdashline & \boldsymbol{u}_1 & & \\ \hdashline & \vdots & & \\ \hdashline & \boldsymbol{u}_{n-1} & & \end{bmatrix}$ be a "formal" determinant. Show that $\bigtimes_{i=1}^{n-1} \boldsymbol{u}_i$ is orthogonal to each $\boldsymbol{u}_i$ and that $\|\bigtimes_{i=1}^{n-1} \boldsymbol{u}_i\|$ is the $(n-1)$-volume of the $(n-1)$-parallelotrope determined by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}$.

Problem 200 (a) Suppose $L : \mathbb{R}^2 \to \mathbb{R}^2$ is linear and R is a bounded region in the plane with area area(R), show that area$(L[R]) = \det[L]$area(R).

(b) Suppose $L : \mathbb{R}^3 \to \mathbb{R}^3$ is linear and S is some solid in $\mathbb{R}^3$ with a volume $\mathrm{Vol}(S)$. Show that $\mathrm{Vol}(L[S]) = \det[L]\,\mathrm{Vol}(S)$.

Problem 201 Consider $A = \begin{bmatrix} 5/3 & 1/3 & 1/3 \\ 1/3 & 11/12 & 5/12 \\ 1/3 & 5/12 & 11/12 \end{bmatrix}$ and let S be the unit ball. Find the volume of $A[S]$. Find the volume of $A[S]$.

Discussion: In the preceding problem $A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$ and so $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$. The unit sphere is

$$S = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \;\middle|\; \left\| \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right\| = 1 \right\}$$

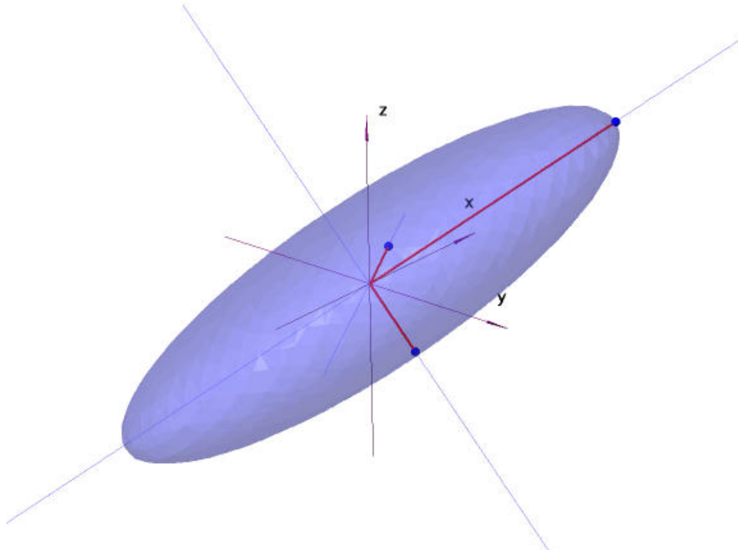
Notice that A is symmetric. We get an equation for $A[S]$ since $\left\| A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right\| = 1$ is equivalent to

$$\begin{aligned}
\left(A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right)^T \left(A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right) &= \begin{bmatrix} u & v & w \end{bmatrix} (A^{-1})^T A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\
&= \begin{bmatrix} u & v & w \end{bmatrix} (A)^{-1} A^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\
&= \begin{bmatrix} u & v & w \end{bmatrix} A^{-2} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\
&= \begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} 1/2 & -1/4 & -1/4 \\ -1/4 & 19/8 & -13/8 \\ -1/4 & -13/8 & 19/8 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\
&= \frac{1}{2} u^2 - \frac{1}{2} uv - \frac{1}{2} uw + \frac{19}{8} v^2 - \frac{13}{4} vw + \frac{19}{8} w^2 = 1
\end{aligned}$$

A is diagonalizable as

$$A = \begin{bmatrix} -1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} -1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}^{-1}$$

The eigenvalues are orthogonal (since A is symmetric) and A acts by stretching in the direction of each eigenvector, the result is the ellipsoid.



Given an $n \times n$ matrix, A, we now have a “simple” check as to whether A is singular, namely

$$\boxed{A \text{ is singular iff } \det(A) = 0.}$$

From this it follows that λ is an eigenvalue of A iff $A - \lambda I$ is singular iff $\det(A - \lambda I) = 0$. The n^{th} degree polynomial $p_A(t) = \det(A - tI)$ is the *characteristic polynomial* of A and

$$\boxed{\lambda \text{ is an eigenvalue of } A \text{ iff } \lambda \text{ is a root of } p_A(t)}$$

Problem 202 For each of the given matrices A, compute the characteristic polynomial, the eigenvalues, and find bases for the associated eigenspaces. If you need to move to $\mathbb{C}$.

(1) $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (reflection about the line $x = y$.)

(2) $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$. (Notice here that $\mathbb{R}^2$ fails to have a basis consisting of eigenvectors.)

(3) $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ (rotation through $90°$).

Problem 203 Show that if A and B are similar, i.e., $A = PBP^{-1}$, then A and B have the same eigenvalues.

A polynomial $p(t)$ with real coefficients factors into a product of powers of the form $x - r$ with r a root and irreducible quadratics $ax^2 + bx + c$. For example

$$p(t) = (t-1)(t-2)^2(t+3)^3(t^2+1)(t^2+t+1)^2$$

The root 2 is said to have algebraic multiplicity 2 and the root -3 algebraic multiplicity 3.

Moving to $\mathbb{C}$, the polynomial factors completely, that is, an n^{th} degree polynomial $p(t)$ factors as $p(t) = c\prod_{i=1}^{k}(t-r_i)^{m_i}$ where r_i are distinct complex roots of $p(t)$ with r_i having multiplicity m_i. For the example above:

$$\begin{aligned} p(t) &= (t-1)(t-2)^2(t+3)^3(t^2+1)(t^2+t+1)^2 \\ &= (t-1)(t-2)^2(t+3)^3(t-i)(t+i)\left(t-\left(\tfrac{-1}{2}-\tfrac{3}{2}i\right)\right)^2\left(t-\left(\tfrac{-1}{2}+\tfrac{3}{2}i\right)\right)^2 \end{aligned}$$

If the polynomial $p(t)$ has real coefficients, then the complex roots appear in complex conjugate pairs, each being the solution to an irreducible at^2+bt+c.

The *algebraic multiplicity* of an eigenvalue, λ, is the multiplicity of the root λ in $p_A(t)$.

For $A \in M_{nn}(\mathbb{C})$ and factoring in $\mathbb{C}$ we get $p_A(t) = \prod_{i=1}^{n}(\lambda_i - t)$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues where each eigenvalue λ is repeated according to its algebraic multiplicity. Notice that

$$\begin{aligned} p_A(t) &= (-t)^n + \left(\sum_{i=1}^{n}\lambda_i\right)(-t)^{n-1} + \cdots + \prod_{i=1}^{n}\lambda_i \\ &= (-t)^n + \operatorname{trace}(\Lambda)(-t)^{n-1} + \cdots + \det(\Lambda) \\ &= (-t)^n + \operatorname{trace}(A)(-t)^{n-1} + \cdots + \det(A) \end{aligned}$$

To see this we need to notice that is $\lambda_1, \ldots, \lambda_n$ is a listing of eigenvalues with repetition, then

$$\boxed{\operatorname{trace}(A) = \operatorname{trace}(\Lambda) = \sum_{i=1}^{n}\lambda_i.}$$

We will verify this for diagonalizable matrices, Jordan canonical form can be used to generalize this to any matrix A. Suppose $A = S\Lambda S^{-1}$ where S is a matrix of eigenvectors. The key lies in

$$\operatorname{trace}(AB) = \operatorname{trace}(BA)$$

Problem 204 Prove this.

So $\operatorname{trace}(S\Lambda S^{-1}) = \operatorname{trace}(S^{-1}(S\Lambda)) = \operatorname{trace}((SS^{-1})\Lambda) = \operatorname{trace}(I\Lambda) = \operatorname{trace}(\Lambda)$. In fact the problem shows that if A and B are similar matrices, then $\operatorname{trace}(A) = \operatorname{trace}(B)$.

Problem 205 Show that for A, B $n \times n$ matrices, $AB - BA \neq I$.

For λ an eigenvalue of A

$$\boxed{\text{geometric multiplicity of } \lambda \leq \text{ algebraic multiplicity of } \lambda}$$

A is diagonalizable precisely when equality holds for each eigenvalue.

Problem 206 Show that the geometric multiplicity of an eigenvalue λ is $\leq$ the algebraic multiplicity. Hint: Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ be a basis for E_λ, where λ is an eigenvalue of A. Expand this set of vectors to a basis $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$ for V. Consider $A = PBP^{-1}$ where $P = \left[\boldsymbol{v}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_k \,\vdots\, \boldsymbol{v}_{k+1} \,\vdots\, \cdots \,\vdots\, \boldsymbol{v}_n\right]$ is the matrix which represents the change of basis from $\mathcal{B}$ coordinates to standard basis coordinates. Notice that $p_A(t) = p_B(t)$.

Problem 207 Show that if

$$A = \left[\begin{array}{c:c} B & C \\ O & D \end{array}\right]$$

is a block triangular matrix with each B and E square, then $\det(A) = \det(B)\det(D)$ and $p_A(t) = p_B(t)p_D(t)$.

6.2 Diagonalization of Hermitian Matrices

The adjoint of A satisfies

$$\langle A\boldsymbol{x}|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|A^*\boldsymbol{y}\rangle,$$

since

$$\langle A\boldsymbol{x}|\boldsymbol{y}\rangle = \boldsymbol{y}^*(A\boldsymbol{x}) = (\boldsymbol{y}^*A^{**})\boldsymbol{x} = (\boldsymbol{y}A^*)^*\boldsymbol{x} = \langle \boldsymbol{x}|A^*\boldsymbol{y}\rangle$$

Conversely, if $B \in M_{nn}(\mathbb{C})$ is such that for all $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\langle A\boldsymbol{x}|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|B\boldsymbol{y}\rangle, \tag{$\dagger$}$$

then $B = A^*$.

This gives a characterization of the adjoint of A, as the unique matrix B satisfying (†).

This characterization of adjoint is not only useful, but also generalizes to an arbitrary linear operator on a finite dimensional inner-product space, namely, if $L : V \to V$ is a linear operator, then the *adjoint of* L, $L^* : V \to V$, is the unique linear operator satisfying:

$$\langle L(\boldsymbol{x})|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|L^*(\boldsymbol{y})\rangle$$

A is *self-adjoint (hermitian)* iff for all $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\langle A\boldsymbol{x}|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|A\boldsymbol{y}\rangle$$

In terms of operators, $L : V \to V$ is self-adjoint iff for all $\boldsymbol{x}, \boldsymbol{y} \in V$,

$$\langle L(\boldsymbol{x})|\boldsymbol{y}\rangle = \langle \boldsymbol{x}|L(\boldsymbol{y})\rangle.$$

Problem 208 For V a finite dimensional inner product space and $L \in \mathcal{L}(V)$, show that λ is an eigenvalue of L iff λ^* is an eigenvalue of L^*. Generally you can not say too much about the associated eigenvectors unless L is normal. (See Problem 214.)

Problem 209 Show that if L is self-adjoint, then every eigenvalue is real.

As a consequence we have that if $A \in M_{nn}(\mathbb{F})$ is hermitian, then the eigenvalues of A are real. Next we see that if A is a hermitian $n \times n$ matrix, then $\mathbb{F}^n$ has an orthogonal basis of eigenvectors.

Show by induction on n that if A is a hermitian $n \times n$ matrix, then A has an orthonormal basis of eigenvectors.

Let λ be an eigenvalue of A and let $\boldsymbol{u}$ be an eigenvector. Let $U = \operatorname{Span}\{\boldsymbol{u}\}$. Notice that for $\boldsymbol{v} \in U^{\perp}$,

$$\langle A\boldsymbol{v}|\boldsymbol{u}\rangle = \langle \boldsymbol{v}|A\boldsymbol{u}\rangle = \langle \boldsymbol{v}|\lambda\boldsymbol{u}\rangle = \lambda\langle \boldsymbol{v}|\boldsymbol{u}\rangle = 0,$$

so $A\boldsymbol{v} \in U^{\perp}$. In other words, $U^{\perp}$ is A-*invariant.*

Consider the linear function $B : U^{\perp} \to U^{\perp}$ given by $B\boldsymbol{v} = A\boldsymbol{v}$. Notice that B remains self adjoint since

$$\langle B\boldsymbol{v}|\boldsymbol{w}\rangle = \langle A\boldsymbol{v}|\boldsymbol{w}\rangle = \langle \boldsymbol{v}|A\boldsymbol{w}\rangle = \langle \boldsymbol{v}|B\boldsymbol{w}\rangle$$

Since $\dim(U^{\perp}) = n - 1$ we know by induction that $U^{\perp}$ has an orthogonal basis of B-eigenvectors which are clearly also A eigenvectors.

Recall that a matrix U is *unitary* if $UU^* = I$ so that $U^{-1} = U^*$ (see Problem 143).

Putting these together we get the *Spectral Decomposition*:

For A a symmetric $n \times n$ matrix, there is a unitary matrix U and a diagonal matrix Λ such that

$$A = U\Lambda U^T$$

If the unitary matrix is the orthogonal matrix of eigenvectors say $U = \left[\, \boldsymbol{u}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{u}_n \,\right]$ where $\boldsymbol{u}_i$ is an eigenvector for λ_i, then the above decomposition gives the following which is also refered to as a spectral decomposition of A:

$$\boxed{A = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T}$$

So

$$A\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i (\boldsymbol{u}_i \boldsymbol{u}_i^T)\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i \operatorname{proj}_{\boldsymbol{u}_i}(\boldsymbol{x}) = \sum_{i=1}^{k} P_{E_i}\boldsymbol{x}$$

where $\mathbb{F}^n = E_1 \oplus \cdots \oplus E_k$ is the decomposition of $\mathbb{F}^n$ into eigenspaces associated to distinct eigenvalues.

This last form of the Spectral decomposition can be written as

$$\boxed{A = \sum_{i=1}^{n} \lambda_i (\boldsymbol{u}_i \boldsymbol{u}_i^T) = \sum_{i=1}^{k} \lambda_i P_{E_i}}$$

Problem 210 Find the spectral decomposition for A and A^3 for each of the following. Let $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, write down $\boldsymbol{b}$ as the sum of its projections into the eigenspaces and write down the distance of $\boldsymbol{b}$ from each eigenspace.

(1) $A = \begin{bmatrix} 4/9 & 4/9 & 1/3 & 2/9 \\ 4/9 & 16/9 & 0 & -4/9 \\ 1/3 & 0 & 11/6 & 1/3 \\ 2/9 & -4/9 & 1/3 & 4/9 \end{bmatrix}$

(2) $A = \begin{bmatrix} {}^{17}\!/_{9} & -{}^{2}\!/_{9} & -{}^{2}\!/_{9} & 0 \\ -{}^{2}\!/_{9} & {}^{14}\!/_{9} & -{}^{4}\!/_{9} & 0 \\ -{}^{2}\!/_{9} & -{}^{4}\!/_{9} & {}^{14}\!/_{9} & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

Problem 211 Assume A and B are diagonalizable, give an example and explain what the problem is with:

> If α is an eigenvalue of A and β is an eigenvalue for B, then $\alpha\beta$ is an eigenvalue for AB

> If α is an eigenvalue of A and β is an eigenvalue for B, then $\alpha + \beta$ is an eigenvalue for $A + B$

Problem 212 Show that if A and B are diagonalizable, then the following are equivalent:

- $AB = BA$ (the matrices commute)
- A and B have a common basis of eigenvectors. (A and B do not have the same eigenvalues.) The fact that A and B share a common basis of eigenvectors can be stated as $A = S\Lambda_A S^{-1}$ and $B = S\Lambda_B S^{-1}$, that is, they share a common eigenvector matrix. In such a case A and B are *simultaneously diagonalizable.*

In the case that A and B commute, and $E_\lambda^A \cap E_{\lambda'}^B \neq \{\mathbf{0}\}$, then $\lambda\lambda'$ is an eigenvalue for AB.

A linear operator $T \in \mathcal{L}(V)$, or equivalently matrix A in $M_{nn}(\mathbb{F})$, is *normal* iff $TT^* = T^*T$.

Problem 213 $T : V \to V$ is normal iff $\|T\boldsymbol{x}\| = \|T^*\boldsymbol{x}\|$ for all $\boldsymbol{x} \in V$.

Problem 214 For $T \in \mathcal{L}(V)$ show that if $TT^* = T^*T$, then $E_\lambda^T = E_{\lambda^*}^{T^*}$, that is $T\boldsymbol{v} = \lambda\boldsymbol{v} \leftrightarrow T^*\boldsymbol{v} = \lambda^*\boldsymbol{v}$. So not only do T and T^* these have the same eigenvectors they have the same eigenspaces.

Next we see that the following are equivalent for $T \in \mathcal{L}(V)$ for V a vector space over $\mathbb{C}$.

- T is unitarily diagonalizable.
- T is normal.

If T is diagonalizable, then $[T] = U\Lambda U^*$ and $[T^*] = U\Lambda^* U^*$ for U unitary, so $U^* = U^{-1}$ and

$$[TT^*] = U\Lambda U^* U\Lambda^* U^* = U\Lambda\Lambda^* U^* = U\Lambda^*\Lambda U^* = U\Lambda U^* U\Lambda^* U^* = [T^*T]$$

so T is normal.

In the other direction suppose T is normal. Let $\boldsymbol{u}$ be an eigenvector for the eigenvalue λ and let $W = \mathrm{Span}(\{\boldsymbol{u}\})^{\perp}$. We know that $\boldsymbol{u}$ is an eigenvector for T^* for λ^*. We just need to show W is both T and T^* invariant and in fact if $S = T|_W$, then $S^* = T^*|_W$ and S is normal. Then induction can be used on $\dim(V)$.

W is both T and T^* invariant: Let $\boldsymbol{w} \in W$, so $\langle \boldsymbol{w}|\boldsymbol{u}\rangle = 0$ we need to show $\langle T(\boldsymbol{w})|\boldsymbol{u}\rangle = 0$. The point is $\langle T(\boldsymbol{w})|\boldsymbol{u}\rangle = \langle \boldsymbol{w}|T^*(\boldsymbol{u})\rangle$ and $T^*(\boldsymbol{u}) \in \mathrm{Span}(\{\boldsymbol{u}\})$, since $\boldsymbol{u}$ is an eigenvector for T^*, so $\langle \boldsymbol{w}|T^*(\boldsymbol{u})\rangle = 0$. This shows that W is T invariant. showing that W is T^* invariant is completely analogous.

If $S = T|_W$, then $S^* = T^*|_W$: This is trivial since

$$\begin{aligned}\langle S(\boldsymbol{w})|\boldsymbol{w}'\rangle &= \langle T|_W(\boldsymbol{w})|\boldsymbol{w}'\rangle = \langle T(\boldsymbol{w})|\boldsymbol{w}\rangle \\ &= \langle \boldsymbol{w}|T^*(\boldsymbol{w}')\rangle = \langle \boldsymbol{w}|T^*|_W(\boldsymbol{w}')\rangle\end{aligned}$$

So $T^*|_W = S^*$.

6.3 Bilinear and quadratic forms

Let V be a vector space. A *bilinear form* is a function $B : V \times V \to \mathbb{F}$, where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$ satisfying

(i) $B(\alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2, \boldsymbol{v}) = \alpha_1 B(\boldsymbol{u}_1, \boldsymbol{v}) + \alpha_2 B(\boldsymbol{u}_2, \boldsymbol{v})$. (Linear in the left coordinate.)

(ii) $B(\boldsymbol{v}, \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2) = \alpha_1^* B(\boldsymbol{v}, \boldsymbol{u}_1) + \alpha_2^* B(\boldsymbol{v}, \boldsymbol{u}_2)$. (Conjugate linear in the right coordinate.)

Notice that if $\mathbb{F} = \mathbb{R}$, then this reduces to linear in both coordinates, hence bi-linear.

6.4 Singular Value Decomposition (SVD)

Here we aim to produce a useful decomposition that applies to an arbitrary matrix A that is similar to diagonalization. One attempt at "diagonalizing" an $m \times n$ matrix A of rank k would be to produce a basis $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$ for $\mathbb{F}^n$ and $\mathcal{C} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_m\}$ where $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a basis for $\mathrm{Ker}(A)^{\perp}$ and $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$ with $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ a basis for $\mathrm{Img}(A)$ and $\{\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_m\}$ a basis for $\mathrm{Img}(A)^{\perp}$.

This can easily be accomplished and we get

$$A = C\Sigma B^{-1} \text{ where } \Sigma = [A]_{C,B} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots \end{bmatrix}$$

In general there is nothing particularly useful about this decomposition. It would be nice to have $\mathcal{B}$ and $\mathcal{C}$ be othonormal bases. We can pick $\mathcal{B}$ to be orthonormal, the problem is in getting $\{A\boldsymbol{v}_1, \dots, A\boldsymbol{v}_k\}$ to be orthogonal.

The trick is to consider A^*A, this is a hermitian positive matrix so all eigenvalues are real and non-negative. List the eigenvalues of A^*A in descending order (with repetition) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, and let $\boldsymbol{v}_i$ be an eigenvector for λ_i. We know $\lambda_i = 0$ for $i > k$ since $\operatorname{rank}(A) = k$. For $1 \leq i, j \leq k$ we have

$$\boldsymbol{v}_i^* A^* A \boldsymbol{v}_j = \langle A\boldsymbol{v}_j | A\boldsymbol{v}_i \rangle = \lambda_j (\boldsymbol{v}_i^* \boldsymbol{v}_j) = \lambda_j \langle \boldsymbol{v}_i | \boldsymbol{v}_j \rangle = 0$$

so $\{A\boldsymbol{v}_1, \dots, A\boldsymbol{v}_k\}$ is an orthogonal basis for $\operatorname{Img}(A)$.

We see that

$$\boldsymbol{v}_i^* A^* A \boldsymbol{v}_i = \langle A\boldsymbol{v}_i | A\boldsymbol{v}_i \rangle = \|A\boldsymbol{v}_i\|^2 = \lambda_i \langle \boldsymbol{v}_i | \boldsymbol{v}_i \rangle = \lambda_i$$

$\langle A\boldsymbol{v}_i | A\boldsymbol{v}_i \rangle = \sqrt{\lambda_i} = \sigma_i$ so set $\boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i$ and hence $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$. We have $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_k > 0$, these are the *singular values of* A.

Extend $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ by adding an orthonormal basis for $\operatorname{Img}(A)^\perp$.

Let

$$V = \left[\begin{array}{c:c:c:c} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{array}\right] \qquad U = \left[\begin{array}{c:c:c:c} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{array}\right]$$

$$\Sigma = \left[\begin{array}{cccc:ccc} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_k & 0 & \cdots & 0 \\ \hdashline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right] = \left[\begin{array}{c:c} \operatorname{diag}(\sigma_1, \dots, \sigma_k) & O \\ \hdashline O & O \end{array}\right]$$

Since V is unitary $V^{-1} = V^*$ and the following is the singular value decomposition:

$$\boxed{A = U\Sigma V^* = \left[\begin{array}{c:c:c:c} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_k \end{array}\right] \operatorname{diag}(\sigma_1, \dots, \sigma_k) \left[\begin{array}{c} \boldsymbol{v}_1^* \\ \hdashline \boldsymbol{v}_2^* \\ \hdashline \vdots \\ \hdashline \boldsymbol{v}_k^* \end{array}\right] = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*}$$

Note for A a $m \times n$ matrix, U is $m \times m$, Σ is $m \times n$, and V^* is $n \times n$. In the above it is clear that we can ignore the “0” part.