Lecture : Linear Algebra Notes

Fall 2020

Lecturer: Dan Calderone

1 Vector Products and Matrix Multiplication

Inner products

General notation: $\langle y, x \rangle$

   Specific inner products:

- Vectors in $\mathbb{R}^n$: $\langle y, x \rangle = y \cdot x = y^T x = \sum_{i=1}^n y_i x_i$

- Vectors in $\mathbb{C}^n$: $\langle y, x \rangle = y^* x = \sum_{i=1}^n y_i^* x_i$

- Integrable functions on $f : [0,1] \to \mathbb{C}^n$: $\langle f, g \rangle = \int_{[0,1]} f^*(t) g(t) \, dt$

One of the fundamental uses of an inner product is to compute the *2-norm* or *length* of a vector by taking an inner product of vector with itself. $|x|_2 = \sqrt{\langle x, x \rangle}$. More generally, inner products tell you how much two vectors *line up with each other*. Along these lines, we have the identity

$$\sqrt{\langle x, x \rangle} = y^T x = |y||x| \cos(\theta) \tag{1}$$

where θ is the angle between x and y. A way to see this directly is to apply the law of cosines to $|x - y|^2$
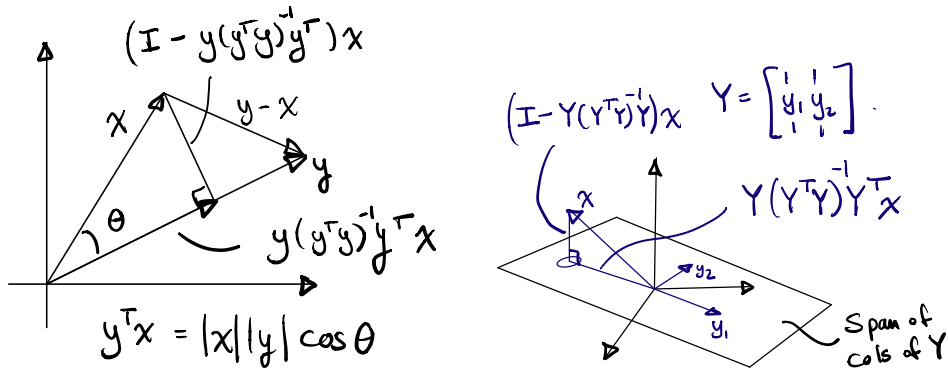
$$(x - y)^T (x - y) = x^T x + y^T y - 2x^T y = |x|^2 + |y|^2 - 2|x||y| \cos(\theta) \tag{2}$$

When $y^T x = 0$, $cos(\theta) = 0$ and the angle between the two vectors is either 90° and -90° and the vectors are *perpendicular* or *orthogonal*. If y is a *unit vector*, ie. $|y| = 1$, then $y^T x = |x| cos(\theta)$, ie. $y^T x$ is the amount of x in the direction of y. If we then multiply this quantity by the unit vector y again, we get the component of x in the y-direction or the *projection of x onto y*, $\text{proj}_y x$. If y is not a unit vector, we can use the unit vector $y/|y|$. This leads to the general formula for a 1-dimensional projection matrix

$$\text{proj}_y x = \frac{1}{|y|^2} y y^T x = y(y^T y)^{-1} y^T x \tag{3}$$

More generally, if we want to project x onto a large subspace spanned by the columns of Y, we can compute

$$\text{proj}_Y x = Y(Y^T Y)^{-1} Y^T x \tag{4}$$

$$\left(I - y(y^T y)^{-1} y^T\right) x$$

x

$y - x$

y

$y(y^T y)^{-1} y^T x$

$y^T x = |x||y| \cos \theta$

$\left(I - Y(Y^T Y)^{-1} Y\right) x \qquad Y = \begin{bmatrix} | & | \\ y_1 & y_2 \\ | & | \end{bmatrix}$

x

$Y(Y^T Y)^{-1} Y^T x$

y_2

y_1

Span of cols of Y

Outer Products

The *outer product* of x and y is given by

$$xy^T = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_n y_1 & \cdots & x_n y_n \end{bmatrix} \tag{5}$$

Outer products are clearly rank-1 and are sometimes called *dyads*. Note that a 1-dimensional projection matrix is the outer product of a unit vector with itself.

Matrix Inner Products

Let $X, Y \in \mathbb{R}^{n \times m}$. The inner product of two matrices is

$$\sum_i \sum_j X_{ij} Y_{ij} = \text{Tr}(Y^T X) \tag{6}$$

where the trace operator $\text{Tr}(\cdot)$ is the sum of the diagonal elements. The Frobenius-norm of a matrix is equivalent to the vector two norm $|X|_F = \sqrt{\text{Tr}(X^T X)}$.

Norms

Properties of Norms

For a vector space $\mathcal{V}$ over a field $\mathcal{F}$, a **norm** is a nonnegative-valued function $\| \cdot \| : \mathcal{V} \to \mathbb{R}$.
For all $a \in \mathcal{F}$ and all $v, u \in \mathcal{V}$

Subadditivity/triangle inequality:	$\|u + v\| \leq \|u\| + \|v\|$
Absolute homogeneity:	$\|av\| = |a| \|v\|$
Nonnegativity:	$\|v\| \geq 0$
Zero vector:	if $\|v\| = 0$, then $v = 0$

2

For convenience from here on, we will use $|\cdot|$ for both absolute values and norms.
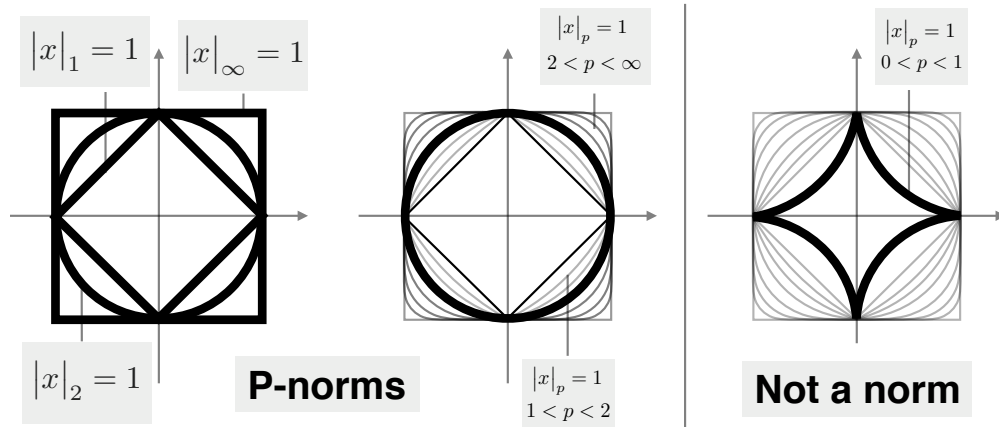
Vector Norms

$$\text{p-norm:} \qquad \left|x\right|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}, \qquad 1 \le p \le \infty$$

$$\text{2-norm:} \qquad \left|x\right|_2 = \left(\sum_i |x_i|^2\right)^{\frac{1}{2}}$$

$$\text{1-norm:} \qquad \left|x\right|_1 = \left(\sum_i |x_i|\right)^{1}$$

$$\infty\text{-norm:} \qquad \left|x\right|_\infty = \lim_{p\to\infty}\left(\sum_i |x_i|^p\right)^{\frac{1}{p}} = \max_i |x_i|$$

P-norms

Not a norm

Norm balls in $\mathbb{R}^3$
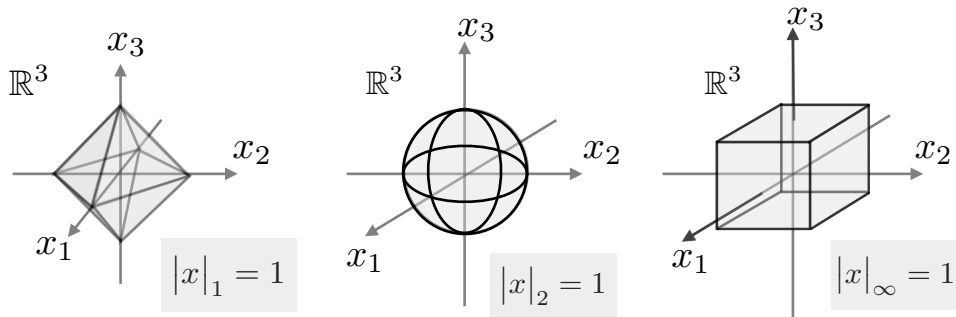
Matrix Norms

Norms for matrices either think of the matrix as a reshaped vector (**element-wise norms**) or as an operator on vector spaces. Norms that treat matrices as operators are called **induced norms**.

Element-wise Matrix Norms

An element-wise matrix 2-norm is called the **Frobenius norm**, $\left| \cdot \right|_{\mathrm{F}}$. For $A \in \mathbb{R}^{m \times n}$

$$\left| A \right|_{\mathrm{F}} = \sum_{ij} \left| A_{ij} \right|^2 = \left(\mathrm{Tr}(A^*A) \right)^{\frac{1}{2}}$$

Note that considering the SVD of $A \in \mathbb{R}^{m \times n}$ (see later on)

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*, \qquad \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix}$$

and applying properties of traces (see later on), we get $\left| A \right|_{\mathrm{F}} = \left| \mathrm{diag}(\Sigma) \right|_2$, ie. the Frobenius norm is the 2-norm applied to a vector of the singular values.

$$\begin{aligned}
\left| A \right|_{\mathrm{F}} &= \left(\sum_{ij} \left| A_{ij} \right|^2 \right)^{\frac{1}{2}} \\
&= \left(\mathrm{Tr}(A^*A) \right)^{\frac{1}{2}} \\
&= \left(\mathrm{Tr}\left(V \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} U^*U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^* \right) \right)^{\frac{1}{2}} \\
&= \left(\mathrm{Tr}\left(\begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V^*V \right) \right)^{\frac{1}{2}} = \left(\sum_i \sigma_i^2 \right)^{\frac{1}{2}}
\end{aligned}$$

Induced Matrix Norms

Induced matrix norms intuitively measure how much a matrix increases (or decreases) the size of vectors it acts on. The induced p, q-norm of $A \in \mathbb{R}^{m \times n}$ gives the maximum q-norm of a vector $\left| Ax \right|_{\beta}$ where x is chosen from the unit ball of the p-norm.

$$\left| A \right|_{p,q} = \max_{\left| x \right|_p = 1} \left| Ax \right|_q$$

or, equivalently.

$$|A|_{p,q} = \max_{x \neq 0} \frac{|Ax|_q}{|x|_p}$$

Sometimes we use $|\cdot|_p$ to refer to the induced p, p-norm. Some specific induced norm examples (again with SVD given above).

$$|A|_2 = |A|_{2,2} = \max_{|x|_2=1} |Ax|_2$$

$$= \max_{|x|_2=1} (x^* A^* A x)^{\frac{1}{2}}$$

$$= \max_{|x|_2=1} \left(x^* V \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V^* x \right)^{\frac{1}{2}} = \sigma_{\max}$$

Block Matrix Multiplication

Consider a matrix $A \in \mathbb{R}^{m \times n}$ divided up into elements, columns, and rows

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & \cdots & | \\ A_{:1} & & A_{:n} \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} - & A_{1:} & - \\ & \vdots & \\ - & A_{n:} & - \end{bmatrix} \tag{7}$$

where we use the Matlab inspired notation $A_{:j}$ and $A_{i:}$ to represent the ith row and jth column of A respectively. We can define multiplying A by a vector x as

$$Ax = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} \tag{8}$$

$$= \begin{bmatrix} | \\ A_{:1} \\ | \end{bmatrix} x_1 + \cdots + \begin{bmatrix} | \\ A_{:n} \\ | \end{bmatrix} x_n = \begin{bmatrix} [\, -A_{1:}- \,]x \\ \vdots \\ [\, -A_{m:}- \,]x \end{bmatrix} \tag{9}$$

Note that we can interpret Ax as x selecting a particular linear combination of the columns of A. The *range of A* is the span of the columns of A, ie. the set of vectors $y \in \mathbb{R}^m$ that can be reached by selecting a suitable x, $y = Ax$. Alternatively, we can interpret Ax as taking the inner product between x with each of the rows of A. The *nullspace of A* is the set of vectors $x \in \mathbb{R}^n$ such that $Ax = 0$ or the set of vectors that are orthogonal to each of the rows of A.

We now consider multiplying two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times k}$. Note that the inner dimensions must match.

$$AB = \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{1n} & \cdots & a_{11}b_{1k} + \cdots + a_{1n}b_{nk} \\ \vdots & & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{1n} & \cdots & a_{m1}b_{1k} + \cdots + a_{mn}b_{nk} \end{bmatrix} \tag{10}$$

Note that this same formula works if you divide A and B into sub or *block matrices*.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \qquad B = \begin{bmatrix} B_{11} & \cdots & B_{1k} \\ \vdots & & \vdots \\ B_{n1} & \cdots & B_{nk} \end{bmatrix} \tag{11}$$

$$AB = \begin{bmatrix} A_{11}B_{11} + \cdots + A_{1n}B_{1n} & \cdots & A_{11}B_{1k} + \cdots + A_{1n}B_{nk} \\ \vdots & & \vdots \\ A_{m1}B_{11} + \cdots + A_{mn}B_{1n} & \cdots & A_{m1}B_{1k} + \cdots + A_{mn}B_{np} \end{bmatrix} \tag{12}$$

Note that we can divide up A and B into any size sub-blocks as long as the inner dimensions of each appropriate A_{ij} and B_{jk} match. Two specific interesting cases are if we divide up A and B into columns or rows. Dividing A into rows and B into columns gives

$$AB = \begin{bmatrix} - & A_{1:} & - \\ & \vdots & \\ - & A_{n:} & - \end{bmatrix} \begin{bmatrix} | & & | \\ B_{:1} & \cdots & B_{:p} \\ | & & | \end{bmatrix} = \begin{bmatrix} A_{1:}B_{:1} & \cdots & A_{1:}B_{:p} \\ \vdots & & \vdots \\ A_{m:}B_{:1} & \cdots & A_{m:}B_{:p} \end{bmatrix} \tag{13}$$

Here we are taking *the inner products of each row of A with each column of B.* . We could also divide up A into columns and B into rows.

$$AB = \begin{bmatrix} | & & | \\ A_{:1} & \cdots & A_{:n} \\ | & & | \end{bmatrix} \begin{bmatrix} - & B_{1:} & - \\ & \vdots & \\ - & B_{n:} & - \end{bmatrix} = \begin{bmatrix} | \\ A_{:1} \\ | \end{bmatrix} \begin{bmatrix} - & B_{1:} & - \end{bmatrix} + \cdots + \begin{bmatrix} | \\ A_{:n} \\ | \end{bmatrix} \begin{bmatrix} - & B_{n:} & - \end{bmatrix} \tag{14}$$

Note that here, we have computed the *sum of the outer products of the matched columns of A and rows of B*.

We also note the following useful extension of this concept. Consider $A \in \mathbb{R}^{m \times n}$ $M \in \mathbb{R}^{n \times p}$, and $B \in \mathbb{R}^{p \times q}$. Using the inner product form above, we can compute

$$AMB = \begin{bmatrix} A_{1:}MB_{:1} & A_{1:}MB_{:q} \\ \vdots & \vdots \\ A_{m:}MB_{:1} & A_{m:}MB_{:q} \end{bmatrix} \tag{15}$$

It is worth noting that $[AMB]_{ij} = A_{i:}MB_{:j}$ Using the outer product form, we can compute

$$AMB = \sum_k \sum_l \begin{bmatrix} | \\ A_{:k} \\ | \end{bmatrix} M_{kl} \begin{bmatrix} - & B_{l:} & - \end{bmatrix} \tag{16}$$

Note that M_{kl} gives the scaling factor for the dyad $A_{:k}B_{l:}$. In (14), we have taken M to be the identity. Some other common and useful examples of block matrix multiplication are given by

$$AB = A \begin{bmatrix} B_1 & \cdots & B_k \end{bmatrix} = \begin{bmatrix} AB_1 & \cdots & AB_k \end{bmatrix} \tag{17}$$

Note in this example, if each B_j is a column, we can think of the matrix A as transforming each column separately.

$$AB = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} B = \begin{bmatrix} A_1 B \\ \vdots \\ A_n B \end{bmatrix} \tag{18}$$

$$AB = \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = A_1 B_1 + \cdots + A_n B_n \tag{19}$$

$$AB = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \begin{bmatrix} B_1 & \cdots & B_k \end{bmatrix} = \begin{bmatrix} A_1 B_1 & \cdots & A_1 B_k \\ \vdots & & \vdots \\ A_m B_1 & \cdots & A_m B_k \end{bmatrix} \tag{20}$$

2   Inner Products

Vectors contain both the notion of length and direction. While magnitude is a property that vectors share with regular numbers, direction is a uniquely vector property.

   As such we can talk about the relative magnitude of two vectors and we can also talk about whether or not they point in similar directions, opposite directions or whether they are "perpendicular" or "orthogonal" to each other. The "inner product" or "dot product" of two vectors is at the heart of this direction comparison.

Definition

The inner product, denoted by $\langle \cdot, \cdot \rangle$, between two vectors $x, y \in \mathbb{R}^n$ is given by

$$\langle y, x \rangle = \sum_i y_i x_i = y^T x = \|y\|_2 \|x\|_2 \cos(\theta)$$

The expression $y^T x$ uses matrix multiplication notation; y^T is a row vector and x as a column vector. (This expression is perhaps the most clean and useful algebraically.) The final expression $\|y\|\|x\|_2 \cos(\theta)$ is the geometric definition of an inner product that we will consider more below. $\|y\|_2$ and $\|x\|_2$ are the magnitudes of x and y and θ is the angle between the two vectors.

Geometry of Inner Products

The geometry of the inner product can be seen by considering the law of cosines (an extension of the Pythagorean theorem) detailed in the image below.

Note that the law of cosines is an extension of the Pythagorean theorem in that it gives a correction term for when $\theta \neq \pi/2$. This correction term is closely related to the inner product between the vectors that form the sides of the triangle. Consider the norm (squared) of the vector difference $x - y$

$$(x - y)^T(x - y) = x^T x + y^T y - 2y^T x$$

The law of cosines gives that the last term can be expressed as

$$2y^T x = 2\|y\|_2 \|x\|_2 \cos(\theta)$$

which gives the geometric interpretation of the inner product. Note from this definition if two vectors x and y point in similar directions (small θ) then $y^T x$ will be larger; if they are perpendicular ($\theta = \pi/2$) then $y^T x = 0$ and if they point in opposite directions (θ closer to π) then $y^T x$ will be negative. The inner product is closely related to length of one vector after it is projected onto another. Specifically if y is a unit vector, then $y^T x$ is exactly the length of x after it is projected onto y as shown in the image below.

Visualizing Scalar Multiplication

To better visualize the geometry of the inner product, we first consider scalar multiplication of two real numbers in the following way. For the product of any two real numbers $x, y \in \mathbb{R}$, we can think of x as a step-size and y as how many steps we take (not necessarily an integer). On a number line, We can think of this as re-defining a unit for y as x. Visually this is equivalent to dragging the value of 1 on the y number line to the value of x and allowing y to be stretched as well.

As 1 goes to x, y will go to xy.

If x is greater than 1, y gets stretched away from the origin; if x is less than one then y gets shrunk toward the origin. If the sign of x is negative then y also gets flipped to the otherside of the numberline. The value $x = 1$ is special in that it leaves y unchanged and $x = -1$ leaves the magnitude of y unchanged but just flips the side of the numberline. We can illustrate this property in the figure below.
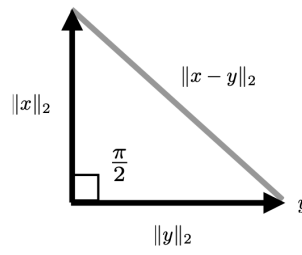
Visualizing Inner Products

Unlike scalars, vectors contain both the notion of length and direction. If a scalar has some notion of direction it is simply a binary value ± 1, ie. is the number positive or negative. This "direction" determines whether or not multiplying by this number flips the direction of another number. Inner products can be thought of as expanding this binary value of ± 1 to a full 360° of relative orientation. Rather than simply flipping the sign of the vector product based on the directions of the vectors, inner product compares their relative direction and then multiplies by a number in the interval $[-1, 1]$ given by $\cos(\theta)$

Law of Cosines (General)
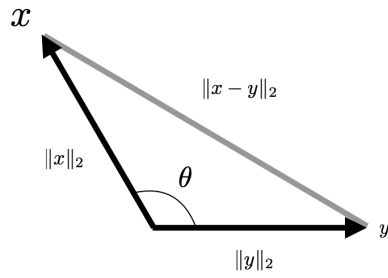
$$\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\|x\|_2\|y\|_2\cos\theta$$

Pythagorean Thm $\left(\theta = \frac{\pi}{2}\right)$

$$\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

$$\mathrm{proj}_y x = \|x\|_2 \cos\theta$$
$$= \frac{1}{\|y\|_2} y^T x$$
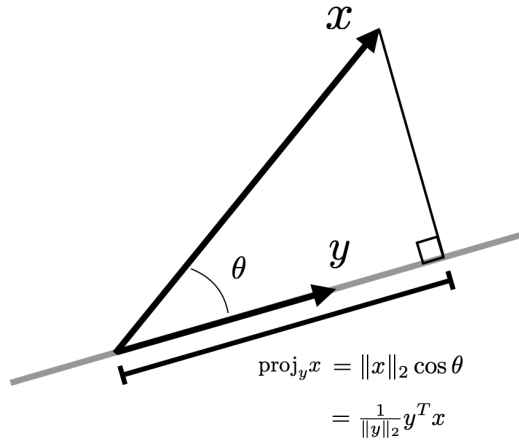
Law of Cosines (General)

$$\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\|x\|_2\|y\|_2\cos\theta$$

Pythagorean Thm $\left(\theta = \frac{\pi}{2}\right)$
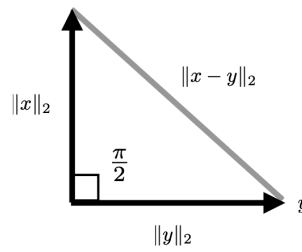

$$\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

9
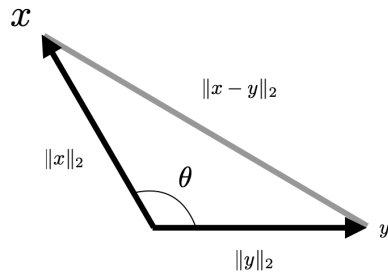
To adjust the visualization of scalar multiplication above to apply to inner products we first note that a unit step for y now includes a direction as well as a length. When taking the inner product rather than redefining the unit vector in the y direction as x, we can find the vector in the y direction, v, that would project to the unit vector in the x direction and redefine that vector as x. If we let, y get stetched in the same way, the resulting length of the stretched y will be $y^T x$. This is visualized in the diagram below. Note that when we move v to x, y moves along a parallel line.

The vector v can be thought of as how far we have to go in the y-direction to get to one unit in the x direction. If x and y point in the same direction, then v just has length one (and this picture reduces to the scalar picture) but the more x and y point in different directions the larger the vector v gets, ie. the farther you have to go in the y direction to move one unit in the x direction. For large v the action of stretching v to x actually ends up shrinking y and so $y^T x$ becomes small. (The diagrams below are dense and worth considering slowly.)

Several other special cases are worthy of note. If x is just a unit vector this operation ends up just giving the length of y projected onto the direction of x. This idea of projection is at the heart of the difference between inner products and scalar products. Scalar products just scale the magnitude of the thing they multiply; inner products both scale and project the things they multiply. Another special case is when y and x are perpendicular. Here the visualization above reaches a limiting case where the vector v shoots off to infinite. Dragging v to x then shrinks y to 0. Intuitively since y does not point in the direction of x at all, we have to go out to infinite to move one unit in the x direction.

The algebra of this visualization is the following. v is the vector in the y-direction such that $v^T \frac{x}{\|x\|_2} = \|v\|_2 \cos\theta = 1$. Since v and y point in the same direction, we can define $y = \beta v$, ie. y is just a scaled version of v. We then get that

$$y^T x = \beta v^T \frac{x}{\|x\|_2} \|x\|_2 = \beta \|x\|_2$$

ie. the same scaling (β) that scales v to y also scales x to a vector with length $y^T x$. The geometric interpretation above is just this visualized using similar triangles.

ORTHOGONALITY

One of the fundamental notions associated with inner products is the idea of orthogonality. Two vectors are orthogonal if they are perpendicular, ie. the angle between them is $\theta = \frac{\pi}{2}$. Intuitivley, if two vectors are orthogonal they represent entirely separate directions that do not affect each other, ie. the projection of one vector onto the other is 0. For a vector $x \in \mathbb{R}^n$, the set of vectors orthogonal to x has dimension $n - 1$. Subspaces are often defined as being orthogonal to some vector or set of vectors (see the discussion of nullspaces).

Inner Products and Quadratic Forms

Note: this section is better understood with background in positive definite matrices.

We can generalize the Euclidean inner product with a positive definite symmetric matrix $P \succ 0$, $P = P^T \in \mathbb{R}^{n \times n}$ to get the P-inner product defined as

Euclidean Inner Product: $y^T x$

$x, y \in \mathbb{R}^2$

Drag v to x...

Same stretching (parallel)
drags y to vector
with length $y^T x$

Point in y-direction
that projects to
the unit vector
in x-direction

y

v

x

$y^T x$

unit ball
$x^T x = 1$

Note: for higher dimensions, the same
picture works within the plane spanned
by the two vectors.

y

x

$y^T x$

$x, y \in \mathbb{R}^3$

$\|x\|_2 < 1$

$\|x\|_2 = 1$

$\|x\|_2 > 1$

y

x

$y^T x$

y

x

$y^T x$

y

x

$y^T x$
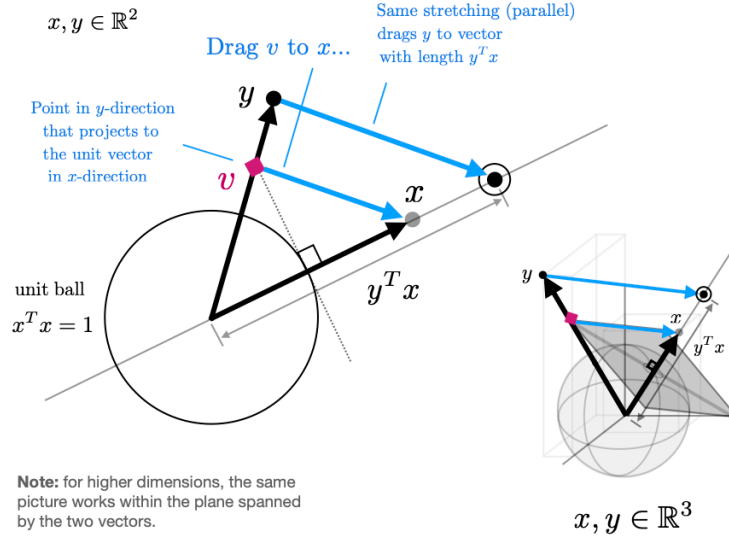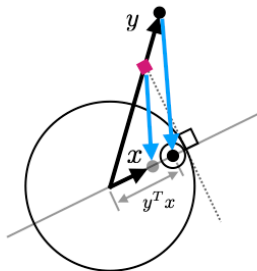
projection

$$\langle y, x \rangle_P = \|x\|_P = y^T P x$$

Rather than a uniform, spherical geometry, this inner product induces an ellipsoidal geometry. One interpretation of this inner product is that each vector x and y is transformed (stretched) using the coordinate transfromations $x' = P^{\frac{1}{2}}x$ and $y' = P^{\frac{1}{2}}y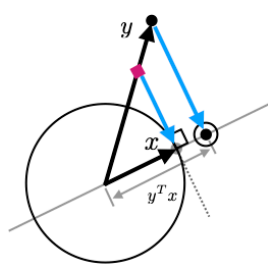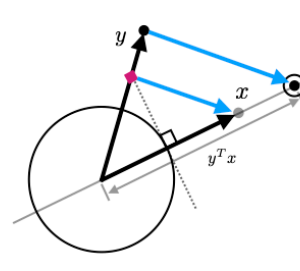$ before the regular Euclidean inner product is applied. The unit ball $x^T P x = 1$ is ellipsoidal rather than spherical in the x-coordinates (but it is spherical in the x'-coordinates). A projection in this ellipsoidal geometry does not follow perpendicular lines but rather lines tangent to the unit ball. Since the ellipsoid changes based on direction, the directions of projection change with direction as well. The visualization of the inner product given above is similar but now the unit vector in the x-direction is shown by the ellisoid and v is the intersection of the y-direction and the tangent space to the ellipsoid where x crosses it. This new geometry is illustrated in the diagram below.

Note that for the P-inner product, orthogonality, ie. $y^T P x = 0$, does not imply that the angle between the two vectors is 90°. Graphically, orthogonality is better understood as one vector being parallel to a tangent vector where the other vector crosses the unit ball. This is illustrated below and discussed more in the section on orthogonality.

COLUMN-COORD GEOMETRY

Another perspective on the inner product $y^T x$ (perhaps closer to the notion of matrix multiplication) is to think of the row vector y^T as a set of columns (each of length) one and then x as a set of coordinates. We can then visualize the inner product in the following way.

3 Derivatives

Vector Derivatives

Derivatives are linear maps that convert perturbations in function arguments into perturbations in the function themselves. Consider $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. $f(x)$ is a scalar. The derivative $\frac{\partial f}{\partial x}$ is the row vector

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

such that

$$\Delta f \approx \frac{\partial f}{\partial x} \Delta x = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} \tag{21}$$

where $\Delta f \in \mathbb{R}$ and $\Delta x \in \mathbb{R}^n$ are perturbations in f and x, respectively. Note that if f is linear, ie. $f(x) = b^\top x$, then $\frac{\partial f}{\partial x} = b^\top$. Note that the perturbation form in (21) can be useful in computing vector derivatives in tricky situations. For example, suppose $f(x) = x^\top Q x + b^\top x$. In order to compute the derivative, we can perturb each instance of x separately and add up the perturbations.

P-Inner Product: $\quad y^T P x$

Drag v to x...

$x, y \in \mathbb{R}^2$

Point in
y-direction (v)
such that
$v^T P u = 1$

y

Same stretching (parallel)
drags y to vector
with length $y^T P x$

v

isocline of
$y^T P x$

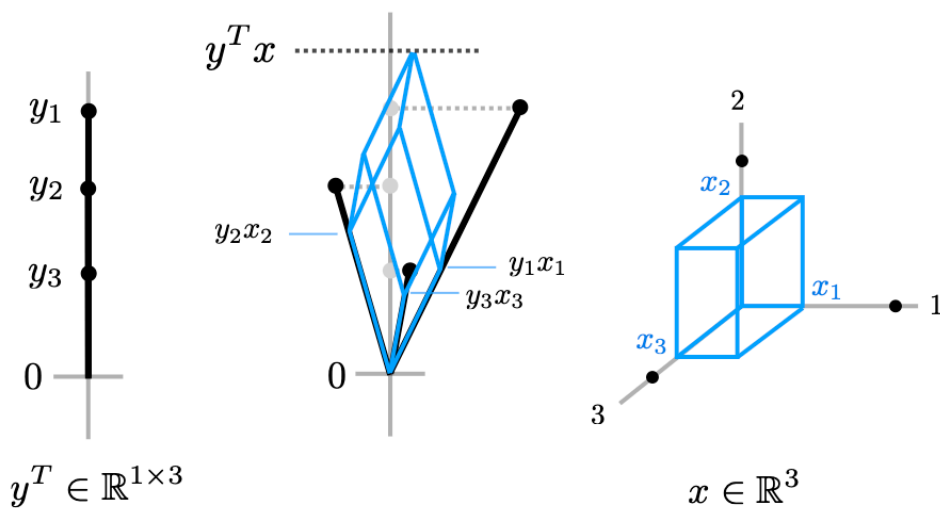**ellipsoidal**

unit ball
$u^T P u = 1$

x

$y^T P x$
(relative to unit ball)

u

$\frac{\partial}{\partial u}(u^T P u) = 2 u^T P$

tangent to unit-ball:

Proof: $\quad v^T P u = 1 = u^T P u$

$\Rightarrow \quad 0 = (v - u)^T P u$

$y^T x$

y_1

y_2

y_3

0

$y_2 x_2$

$y_1 x_1$

$y_3 x_3$

0

2

x_2

x_1

1

x_3

3

$y^T \in \mathbb{R}^{1 \times 3}$

$x \in \mathbb{R}^3$

(The ability to perturb each instance of x separately is called the *product rule*.) Then we rearrange the right hand side (RHS) into the form of (21).

$$\Delta f = \Delta x^\top Q x + x^\top Q \Delta x + b^\top \Delta x \tag{22}$$

Noticing that each of the terms in the RHS is a scalar, we can transpose as necessary.

$$\Delta f = (\Delta x^\top Q x)^\top + x^\top Q \Delta x + b^\top \Delta x \tag{23}$$

$$= \left(x^\top (Q + Q^\top) + b^\top\right)\Delta x \tag{24}$$

$$\Rightarrow \quad \frac{\partial f}{\partial x} = x^\top (Q + Q^\top) + b^\top \tag{25}$$

Now suppose $f(x)$ is a vector valued function $f : \mathbb{R}^n \to \mathbb{R}^n$. The derivative is now an $m \times n$ matrix

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{26}$$

such that

$$\Delta f = \begin{bmatrix} \Delta f_1 \\ \vdots \\ \Delta f_m \end{bmatrix} \approx \frac{\partial f}{\partial x}\Delta x = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} \tag{27}$$

where $\Delta f \in \mathbb{R}^m$ and $\Delta x \in \mathbb{R}^n$. Note that when $\frac{\partial f}{\partial x}$ is a matrix it is referred to as a *Jacobian*.

Now suppose we have a scalar function $f(x)$ and we want to compute its second derivative. Differentiating once gives

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{28}$$

Now treating $\frac{\partial f}{\partial x}$ as a vector valued function, we can compute the second derivative

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{29}$$

The matrix $\frac{\partial^2 f}{\partial x^2}$ is symmetric since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ and is referred to as the *Hessian* of the function $f(x)$. Second derivatives are used to approximate perturbations of first derivatives

$$\Delta \frac{\partial f}{\partial x} \approx \Delta x^\top \frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}^\top \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{30}$$

14

For the quadratic function $f(x) = x^\top Q x + b^\top x$, we can use the perturbative perspective to compute

$$\Delta \frac{\partial f}{\partial x} = \Delta x^\top \frac{\partial^2 f}{\partial x^2} = \Delta x^\top (Q + Q^\top) \Rightarrow \qquad \frac{\partial^2 f}{\partial x^2} = Q + Q^\top \qquad (31)$$

Note that often we write $\frac{\partial^2 f}{\partial x^2} = 2Q$. This is consistent with above formula assuming that $Q = Q^\top$ is symmetric. Any time we consider a quadratic form $x^\top Q x$, we assume that Q is symmetric. The reason for this is that if it's not symmetric, only the symmetric part of it affects the product $x^\top Q x$. Explicitly, write

$$
\begin{aligned}
x^\top Q x &= x^\top \left(\tfrac{1}{2}(Q + Q^\top) + \tfrac{1}{2}(Q - Q^\top) \right) x \\
&= \tfrac{1}{2} x^\top (Q + Q^\top) x + \tfrac{1}{2} x^\top (Q - Q^\top) x \\
&= \tfrac{1}{2} x^\top (Q + Q^\top) x + \underbrace{\tfrac{1}{2} x^\top Q x - \tfrac{1}{2} x^\top Q^\top x}_{=0}
\end{aligned}
$$

The first part of the expansion is the symmetric part of Q. The second part is the skew symmetric part and $x^\top K x = 0$ for any $K = -K^\top$ (K is skew-symmetric).

Using this structure, we also comment on how to express a vector valued Taylor expansion. Up to the quadratic term a Taylor expansion for $f(x)$ around a point x_0 is given by

$$f(x) = f(x_0) + \left.\frac{\partial f}{\partial x}\right|_{x_0} \Delta x + \Delta x^\top \left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0} \Delta x + \cdots \quad \text{where} \quad \Delta x = x - x_0$$

Note how this relates to the perturbation analysis ideas discussed above.

Chain Rule

One important practical tool for taking vector derivatives is the *chain rule*. One of the reasons to be careful about how we arrange vector derivatives, and particularly to write the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ as an $m \times n$ matrix $\frac{\partial f}{\partial x} \in \mathbb{R}^{m \times n}$ is so that it is easy to apply the chain rule consistent with the rules of multiplication. Specifically, consider several functions

$$h(z) : \mathbb{R}^q \to \mathbb{R}^m, \qquad g(y) : \mathbb{R}^p \to \mathbb{R}^q, \qquad f(x) : \mathbb{R}^n \to \mathbb{R}^p$$

The derivatives of each function are matrices

$$\frac{\partial h}{\partial z} \in \mathbb{R}^{m \times q}, \qquad \frac{\partial g}{\partial y} \in \mathbb{R}^{q \times p}, \qquad \frac{\partial f}{\partial x} \in \mathbb{R}^{p \times n}$$

Suppose these functions are now composed together $u(x) = h(g(f(x)))$. The derivative of $u(x)$ with respect to x can then be computed as

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x}\Big(h(g(f(x)))\Big) = \left[\frac{\partial h}{\partial z}\right] \left[\frac{\partial g}{\partial y}\right] \left[\frac{\partial f}{\partial x}\right]$$

15

Carefully note the order of the vector derivative matrices and also how the dimensions of each matrix match up for the matrix multiplication to work. Note also how our perturbation analysis goes through.

$$\Delta u = \frac{\partial u}{\partial x}\Delta x = \left[\frac{\partial h}{\partial z}\right]\left[\frac{\partial g}{\partial y}\right]\underbrace{\left[\frac{\partial f}{\partial x}\right]\Delta x}_{\Delta y}$$
$$\underbrace{\phantom{\left[\frac{\partial h}{\partial z}\right]\left[\frac{\partial g}{\partial y}\right]\left[\frac{\partial f}{\partial x}\right]\Delta x}}_{\Delta z}$$

To be completely accurate we have to be careful to plug in the correct argument to each derivative matrix and thus we should write

$$\frac{\partial u}{\partial x} = \left.\frac{\partial h}{\partial z}\right|_{g(h(x))} \left.\frac{\partial g}{\partial y}\right|_{f(x)} \left.\frac{\partial f}{\partial x}\right|_{x}$$

As an example consider the function $u(x) = e^{-\frac{1}{2}y^\top Q y}$ where $y = Hx$ for $y \in \mathbb{R}^p$ and $H \in \mathbb{R}^{p \times n}$. (This is essentially the equation for a slice of a multivariate Gaussian.). Here we can take

$$h(z) = e^z, \qquad g(y) = -\tfrac{1}{2}y^\top Q y, \qquad f(x) = Hx$$

with derivatives

$$\frac{\partial h}{\partial z} = e^z \in \mathbb{R}^{1 \times 1}, \qquad \frac{\partial g}{\partial y} = -\tfrac{1}{2}y^\top(Q + Q^\top) \in \mathbb{R}^{1 \times p}, \qquad \frac{\partial f}{\partial x} = H \in \mathbb{R}^{p \times n}$$

Plugging in $y = Hx$ and $z = -\frac{1}{2}y^\top Q y$ gives

$$\frac{\partial u}{\partial x} = -e^z\Big|_{-\frac{1}{2}x^\top H^\top Q H x} \tfrac{1}{2}y^\top(Q + Q^\top)\Big|_{Hx} H$$
$$= -\tfrac{1}{2}e^{-\frac{1}{2}x^\top H^\top Q H x}x^\top H^\top\tfrac{1}{2}(Q + Q^\top)H$$
$$= -\tfrac{1}{2}e^{-\frac{1}{2}x^\top H^\top Q H x}x^\top H^\top Q H$$

where in the last line we've assumed Q is symmetric. Note carefully how all the dimensions work out so that the above expression is consistent with the rules of matrix multiplication. Again, the fact that the dimensions work out is not a fluke but rather because we were careful to be consistent with our definition of derivatives and application of the chain rule.

Matrix Derivatives

We now consider taking derivatives of functions $F(X)$ where either the input X or the output F are matrices. The perturbation analysis from above works exactly the same, but these are generally trickier to write down because they are usually higher (more than two) dimensional tensors. The one exception which we will deal with first is when either X or F is simply a scalar.

We start with the case where X is a scalar, $F : \mathbb{R} \to \mathbb{R}^{m \times n}$. In this case, we will usually define

$$\frac{\partial F}{\partial X} = \begin{bmatrix} \frac{\partial F_{11}}{\partial X} & \cdots & \frac{\partial F_{1n}}{\partial X} \\ \vdots & & \vdots \\ \frac{\partial F_{m1}}{\partial X} & \cdots & \frac{\partial F_{mn}}{\partial X} \end{bmatrix} \tag{32}$$

The perturbation analysis can then be written as

$$\Delta F \approx= \frac{\partial F}{\partial X} \Delta X = \begin{bmatrix} \frac{\partial F_{11}}{\partial X} & \cdots & \frac{\partial F_{1n}}{\partial X} \\ \vdots & & \vdots \\ \frac{\partial F_{m1}}{\partial X} & \cdots & \frac{\partial F_{mn}}{\partial X} \end{bmatrix} \Delta X$$

Each element of $\frac{\partial F}{\partial X}$ is simply the scalar derivative of the corresponding element of F with respect to X. $\left[\frac{\partial F}{\partial X}\right]_{ij} = \frac{\partial F_{ij}}{\partial X}$.

We now consider the case where F is a scalar and X is a matrix, $F : \mathbb{R}^{m \times n} \to \mathbb{R}$. In this case,

$$\frac{\partial F}{\partial X} = \begin{bmatrix} \frac{\partial F}{\partial X_{11}} & \cdots & \frac{\partial F}{\partial X_{ij}} \\ \vdots & & \vdots \\ \frac{\partial F}{\partial X_{m1}} & \cdots & \frac{\partial F}{\partial X_{mn}} \end{bmatrix} \tag{33}$$

Note the similarities and differences with (32). The perturbation analysis will involve summing over all elements of $\frac{\partial F}{\partial X}$. We could, for example, write

$$\Delta F = \sum_{ij} \frac{\partial F}{\partial X_{ij}} \Delta X_{ij}$$

However, in many practical problems, scalar functions of matrices are written in terms of quadratic forms or trace operators such as $F(X) = a^\top X b$ with $a \in \mathbb{R}^m$, $b \in \mathbb{R}^n$ or $F(X) = \text{Tr}(C^\top X)$ with $C \in \mathbb{R}^{m \times n}$. It is worth knowing how to deal with these cases specially. Our analysis will leverage properties of the trace operator and Euclidean matrix inner product $\langle C, X \rangle = \text{Tr}(C^\top X)$.

Paralleling the notation of a vector dot product, the basic inner product on the space of matrices is

$$\langle Y, X \rangle = \sum_{ij} Y_{ij} X_{ij} = \text{Tr}(Y^\top X)$$

Here we simply match up the corresponding elements of Y and X and sum over them. One can check that the final expression $\text{Tr}(Y^\top X)$ does exactly this. (In practice, one would not compute the full product $Y^\top X$ in order to calculate this inner product cause only the diagonal is needed, but it is quite useful for analytic purposes.). Using this inner product idea, we can rewrite our perturbation analysis as

$$\Delta F = \sum_{ij} \frac{\partial F}{\partial X_{ij}} \Delta X_{ij} = \left\langle \frac{\partial F}{\partial X}, \Delta X \right\rangle = \text{Tr}\left(\frac{\partial F}{\partial X}^\top \Delta X\right) \tag{34}$$

17

Again, this can be a useful way to think of $\frac{\partial F}{\partial X}$, it is the matrix object that if we take the matrix inner product of it with a perturbation ΔX then we get the perturbation in F, ΔF. The trace expression can also be quite useful because it is often easy to write our function $F(X)$ in a form that looks like the far RHS of (34). We give several examples. The function $F(X) = \text{Tr}(C^\top X)$ is in this form already and we immediately have that

$$F(X) = \text{Tr}(C^\top X) \quad \Longrightarrow \quad \frac{\partial F}{\partial X} = C$$

The function $F(X) = a^\top X b$ is a little trickier, but since it is a scalar value we can put it inside a trace operator without changing it, ie. $F(X) = \text{Tr}(F(X)) = \text{Tr}(a^\top X b)$. (This is always possible for any scalar function F). We can then leverage the cyclic property of traces.

$$\text{Tr}(ABCD) = \text{Tr}(DABC) = \text{Tr}(CDAB) = \text{Tr}(BCDA)$$

assuming that ABC was square in the first place. (It is worth playing around with this formula and convincing yourself that is true as well as seeing how the dimensions of A, B, C, D come into play. The only requirement for this to work is that $ABCD$ is square (and that the dimensions of A, B, C, D are compatible for the original multiplication.) For this reason, trace algebra is actually quite pleasant because you can change the order of matrices in a product (which is not possible when the product is not inside a trace). Returning to our original formula we can write

$$F(X) = a^\top X b = \text{Tr}(a^\top X b) = \text{Tr}(ba^\top X) \quad \Longrightarrow \quad \frac{\partial F}{\partial X} = ab^\top$$

Similarly for $F(X) = \text{Tr}(AXB)$, we can write

$$F(X) = \text{Tr}(AXB) = \text{Tr}(BAX) \quad \Longrightarrow \quad \frac{\partial F}{\partial X} = A^\top B^\top$$

In any practical setting where one is taking derivatives with respect to matrices, being able to use these algebraic tricks involving traces is crucial. Trying to compute out each element of (33) individually and then organize them back into a usable expression is not doable.

4   Projections

Projection onto a Vector

One of the fundamental uses of inner products is to compute projections. A projection of a vector x onto y, which we can denote $\text{proj}_y x$, is the closest vector to x that points precisely in the y direction. We illustrate this in the figure below.

Intuitively one can see that the difference between x and $\text{proj}_y x$ must be orthogonal to y. (Mathematically, this comes from the Pythagorean theorem applied to $x - \text{proj}_y x$). We can also see intuitively that the length of the projection is $||x||_2 \cos\theta$ where θ is the angle between x and y

and can thus be related to the inner product using the geometric definition $y^T x = ||x||_2 ||y||_2 \cos \theta$
If a vector y is a unit vector, than the geometric definition of an inner product gives that the length of the projection is simply $y^T x$. Algebraically, then, the process of computing a projection is can be done by converting y to a unit vector (ie. $\frac{y}{||y||_2}$) taking the inner product with x to get the length of the projection and then multiplying that length by the unit vector again in the y direction. Formulaically, we have that

$$\text{proj}_y x = \left(\frac{1}{||y||_2} y \right) \left(\frac{1}{||y||_2} y^T x \right) = y(y^T y)^{-1} y^T x$$

The second version in the equation above is elegant in that it explicitly writes the projection operator as a matrix that we can then use to project any vector x onto y. We can denote this matrix $\text{proj}_y = y(y^T y)^{-1} y^T$. We will also see that this form is directly extendable to projecting onto a subspace of dimension greater than one.

If we want to get the component of x orthogonal to y we can simply subtract the projection from x. This can be called the projection orthogonal to y.

$$\text{proj}_y^\perp x = x - y(y^T y)^{-1} y^T x = \left[I - y(y^T y)^{-1} y^T \right] x$$

Note that again we have managed to write the projection operation as a matrix which we can denote $\text{proj}_y^\perp = I - y(y^T y)^{-1} y^T$.

It is instructive to check that for any x, $\text{proj}_y x$ and $\text{proj}_y^\perp x$ are orthogonal to each other. We can check this actually independent of x as follows

$$\left(\text{proj}_y x \right)^T \text{proj}_y^\perp x = x^T y(y^T y)^{-1} y^T \left(I - y(y^T y)^{-1} y^T \right) x$$

$$= x^T \left(y(y^T y)^{-1} y^T - y(y^T y)^{-1} y^T y(y^T y)^{-1} y^T \right) x$$

$$= x^T \left(y(y^T y)^{-1} y^T - y(y^T y)^{-1} y^T \right) x = x^T (0) x = 0$$

Note the fact that $\text{proj}_y (\text{proj}_y) = \text{proj}_y$.

Projection onto a Subspace

A similar formulation works if we would like to project x onto a a subspace spanned by multiple vectors given by the columns of a matrix $Y \in \mathbb{R}^{m \times n}$

$$Y = \begin{bmatrix} | & & | \\ Y_1 & \cdots & Y_n \\ | & & | \end{bmatrix}$$

We will assume for this discussion the columns of Y are linearly independent and we will denote this as $\text{proj}_Y x$.

$x \in \mathbb{R}^3$

x

$\mathrm{proj}_y^\perp x = x - \mathrm{proj}_y x$
$= \underbrace{\left[I - y(y^T y)^{-1} y^T \right]}_{\substack{\text{orthogonal projection} \\ \text{matrix}}} x$

2

θ

y

1

$\mathrm{proj}_y x = \dfrac{y}{\|y\|_2} \|x\|_2 \cos\theta$

$= \dfrac{y}{\|y\|_2} \dfrac{y^T x}{\|y\|_2}$

$= \underbrace{\left[y(y^T y)^{-1} y^T \right]}_{\substack{\text{projection} \\ \text{matrix}}} x$

$x \in \mathbb{R}^3$

x

$\mathrm{proj}_Y^\perp x = \underbrace{\left(I - Y(Y^T Y)^{-1} Y^T \right)}_{\substack{\text{orthogonal projection} \\ \text{matrix}}} x$

2

$Y = \begin{bmatrix} | & | \\ Y_1 & Y_2 \\ | & | \end{bmatrix}$

Y_2

$\mathrm{span}\ Y$ ——

Y_1

$\mathrm{proj}_Y x = \underbrace{Y(Y^T Y)^{-1} Y^T x}_{\substack{\text{projection} \\ \text{matrix}}}$

1

3

As in the one dimensional case above, we proceed by converting columns of Y into unit vectors, but for multi-dimensional subspaces there is another subtlety; we must make the columns of Y orthogonal to each other as well. To see this consider a counter-example for two vectors Y_1 and Y_2. A naive approach would be to normalize each vector Y_i, project x onto each and add up the projections according to the formula

$$\text{proj}_Y x = \text{proj}_{Y_1} x + \text{proj}_{Y_2} x$$

This is, in general, an incorrect approach as illustrated in the left figure below Unless, Y_1 and Y_2 are orthogonal to each other the result will not give the desired projection. If, however, the columns of Y are orthonormal as illustrated in the right figure.

The process of orthonormalizing the columns of Y is more complicated than normalizing a single vector. We will discuss this in much more detail in the section on orthonormalization and in the section on shape matrices and polar decomposition. For now we will simply give the following formula. If we take

$$U = Y(Y^T Y)^{-\frac{1}{2}}$$

the columns of U are orthonormal to each other and span the same space as Y. Critically, we can check that the columns of U are orthonormal by taking all the pairwise inner products of each column.

$$U^T U = (Y^T Y)^{-\frac{1}{2}} Y^T Y (Y^T Y)^{-\frac{1}{2}} = I$$

Here we note that the matrix $(Y^T Y)^{\frac{1}{2}}$ is invertible if and only if the columns of Y are linearly independent. We note also that there is a specific relationship between the columns of U and Y that we will discuss more in later sections.

We then can compute the projection of x onto the span of Y by projecting x onto each column of U and summing up.

$$\text{proj}_y x = U_1 U_1^T x + \cdots U_n U_n^T x = U U^T x$$

Note that the second equality is based on the outer product formulation of matrix multiplication. Expanding this out in terms of the original matrix Y now gives

$$\text{proj}_y x = Y(Y^T Y)^{-\frac{1}{2}} (Y^T Y)^{-\frac{1}{2}} Y^T x = Y(Y^T Y)^{-1} Y^T x$$

Here we can denote the projection matrix as $\text{proj}_Y = Y(Y^T Y)^{-1} Y^T$. Note the similarities to the one-dimesional form and specifically the similar role of $(y^T y)^{-1}$ and $(Y^T Y)^{-1}$ in the two formulas.

Similarly, we can compute a projection orthogonal to the span of Y as

$$\text{proj}_Y^\perp x = x - Y(Y^T Y)^{-1} Y^T x = \left(I - Y(Y^T Y)^{-1} Y^T\right) x$$

with projection matrix $\text{proj}_Y^\perp = \left(I - Y(Y^T Y)^{-1} Y^T\right)$. Finally, we can check again algebraically that for any x, $\text{proj}_Y x$ and $\text{proj}_Y^\perp x$ are orthogonal (as desired) regardless of what x is.

$$\begin{aligned}
\left(\text{proj}_Y x\right)^T \text{proj}_Y^\perp x &= x^T Y(Y^T Y)^{-1} Y^T \left(I - Y(Y^T Y)^{-1} Y^T\right) x \\
&= x\left(Y(Y^T Y)^{-1} Y^T - Y(Y^T Y)^{-1} Y^T Y(Y^T Y)^{-1} Y^T\right) x \\
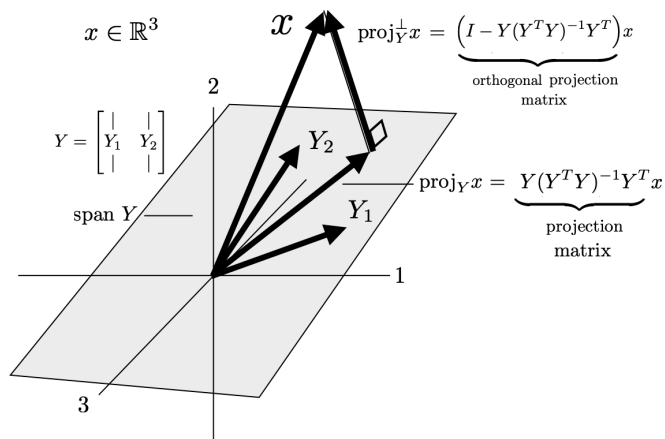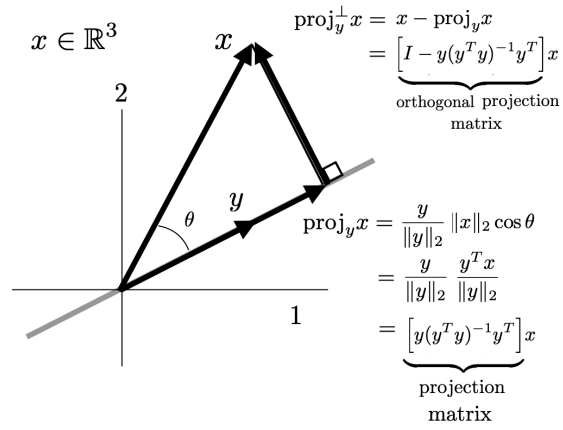&= x\left(Y(Y^T Y)^{-1} Y^T - Y(Y^T Y)^{-1} Y^T\right) x = x^T(0) x = 0
\end{aligned}$$

We now give a detailed visualization of the process discussed above.

More Abstraction: Abstract Projection Definitions

Generally in math, a projection operator is an operator such that applying the operator twice yields the same result as applying it once, ie.

$$(\text{proj}_Y)(\text{proj}_Y) = \text{proj}_Y$$

This is enough to imply the orthogonality conditions we expect above since it implies that

$$\left(\text{proj}_Y x\right)^T \left(I - \text{proj}_Y\right) x = x^T \left(\text{proj}_Y - (\text{proj}_Y)\text{proj}_Y\right) x = x^T \left(\text{proj}_Y - \text{proj}_Y\right) x = x^T (0) x = 0$$

We also can check that this true for for the matrix operators above

$$\text{proj}_Y(\text{proj}_Y) = Y(Y^TY)^{-1}Y^TY(Y^TY)^{-1}Y^T = Y(Y^TY)^{-1}Y^T = \text{proj}_Y$$

and

$$\text{proj}_Y^\perp(\text{proj}_Y^\perp) = \left(I - Y(Y^TY)^{-1}Y^T\right)\left(I - Y(Y^TY)^{-1}Y^T\right)$$
$$= I - 2Y(Y^TY)^{-1}Y^T + Y(Y^TY)^{-1}Y^TY(Y^TY)^{-1}Y^T$$
$$= I - 2Y(Y^TY)^{-1}Y^T + Y(Y^TY)^{-1}Y^T = I - Y(Y^TY)^{-1}Y^T = \text{proj}_Y^\perp$$

5 Complex Numbers

- Complex number: $z \in \mathbb{C}$.

- Cartesian representation: $z = a + bi$

  - Vector-like addition: $z_1 + z_2 = (a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2)i$
  - Norm (length): $|z| = \sqrt{z^* z} = \sqrt{(a - bi)(a + bi)} = \sqrt{a^2 + b^2}$
  - Conjugate: $z^* = \bar{z} = a - bi$.
  - Inverse and Conjugate Inverse:

  $$z^{-1} = \frac{1}{a + bi} = \frac{a - bi}{(a + bi)(a - bi)} = \frac{a}{\sqrt{a^2 + b^2}} + \frac{-b}{\sqrt{a^2 + b^2}} i$$
  $$z^{-*} = \bar{z}^{-1} = \frac{1}{a - bi} = \frac{a + bi}{(a - bi)(a + bi)} = \frac{a}{\sqrt{a^2 + b^2}} + \frac{+b}{\sqrt{a^2 + b^2}} i$$

  - Multiplication: $z_1 z_2 = (a_1 + b_1 i)(a_2 + b_2 i) = a_1 a_2 + (a_1 b_1 + a_2 b_2)i + b_1 b_2$.
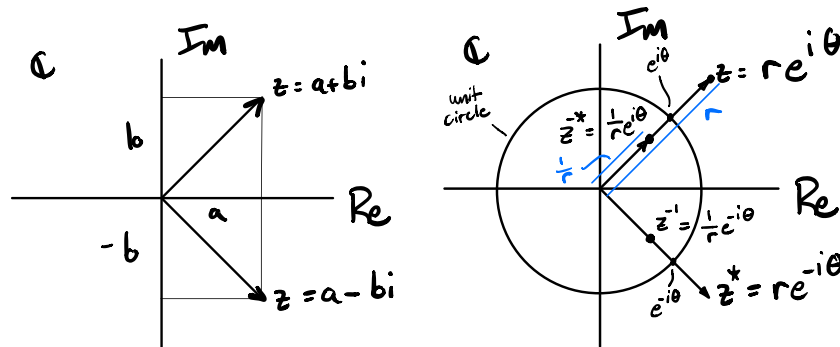
- Polar representation: $z = re^{i\theta}$, $r \geq 0$

    - Relationship to Cartesian representation:

    $$z = a + bi = r\cos(\theta) + r\sin(\theta)i,$$
    $$z = re^{i\theta} = \sqrt{z^*z}\, e^{i\tan^{-1}\left(\frac{a}{b}\right)} = \sqrt{a^2 + b^2}\, e^{i\tan^{-1}\left(\frac{a}{b}\right)}$$

    - Stretching and Rotation:
    The polar represents the stretching and rotational components of a complex number.

    $$z = \underbrace{r}_{\substack{\text{Stretching} \\ \text{by } r}} \underbrace{e^{i\theta}}_{\substack{\text{Rotation} \\ \text{by } \theta}}$$

    - Conjugate: $z^* = \bar{z} = re^{-i\theta}$.
    - Inverse and Conjugate Inverse:

    $$z^{-1} = \tfrac{1}{r}e^{-i\theta}$$
    $$z^{-*} = \bar{z}^{-1} = \tfrac{1}{r}e^{i\theta}$$

    - Multiplication: $z_1 z_2 = r_1 r_2 e^{i\theta_1} e^{i\theta_2} = r_1 r_2 e^{i(\theta_1 + \theta_2)}$

- Roots of Unity:

    - Solutions to the equation: $z^n = 1$.
    - n solutions:

    $$z = e^{i\frac{2\pi k}{n}}, \quad \text{for} \quad k = 0, 1, 2, \ldots, n-2, n-1$$

    - Each solution corresponds to an angle step size $\Delta\theta = \frac{2\pi k}{n}$ and powers of $z = e^{i\frac{2\pi k}{n}}$ represent stepping around the circle. k corresponds to the number of rotations around the unit circle before returning to 1. $k = 0$ is zero rotations, $k = 1$ is one rotation, $k = 2$ is two rotations, etc.
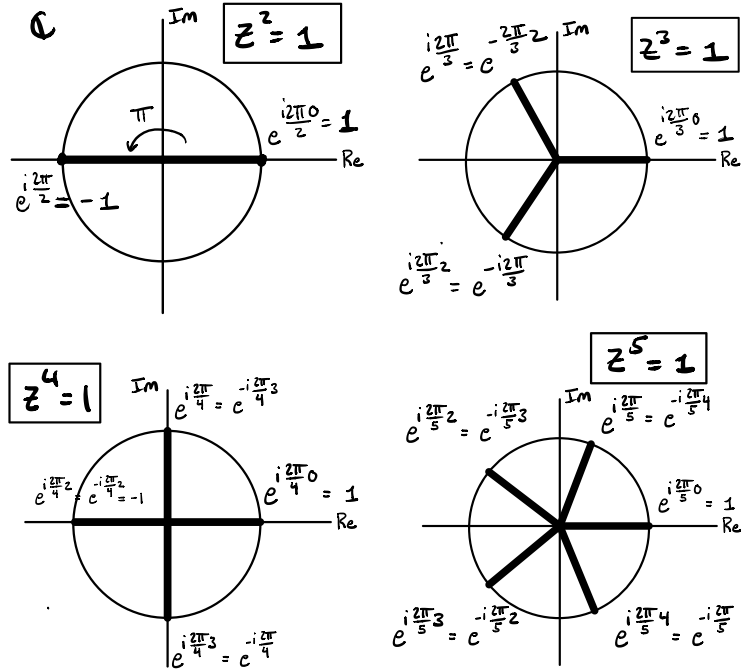   - Alternative enumeration of solutions corresponding to rotating in reverse:

    $$z = e^{i\frac{2\pi(-k')}{n}}, \quad \text{for} \quad k' = n, (n-1), \ldots, 2, 1$$

    by the relationship $k = n - k'$

    $$z = e^{i\frac{2\pi(-k')}{n}} = e^{i\frac{2\pi(-k')}{n}} e^{i\frac{2\pi n}{n}} = e^{i\frac{2\pi(n-k')}{n}} = e^{i\frac{2\pi k}{n}}$$

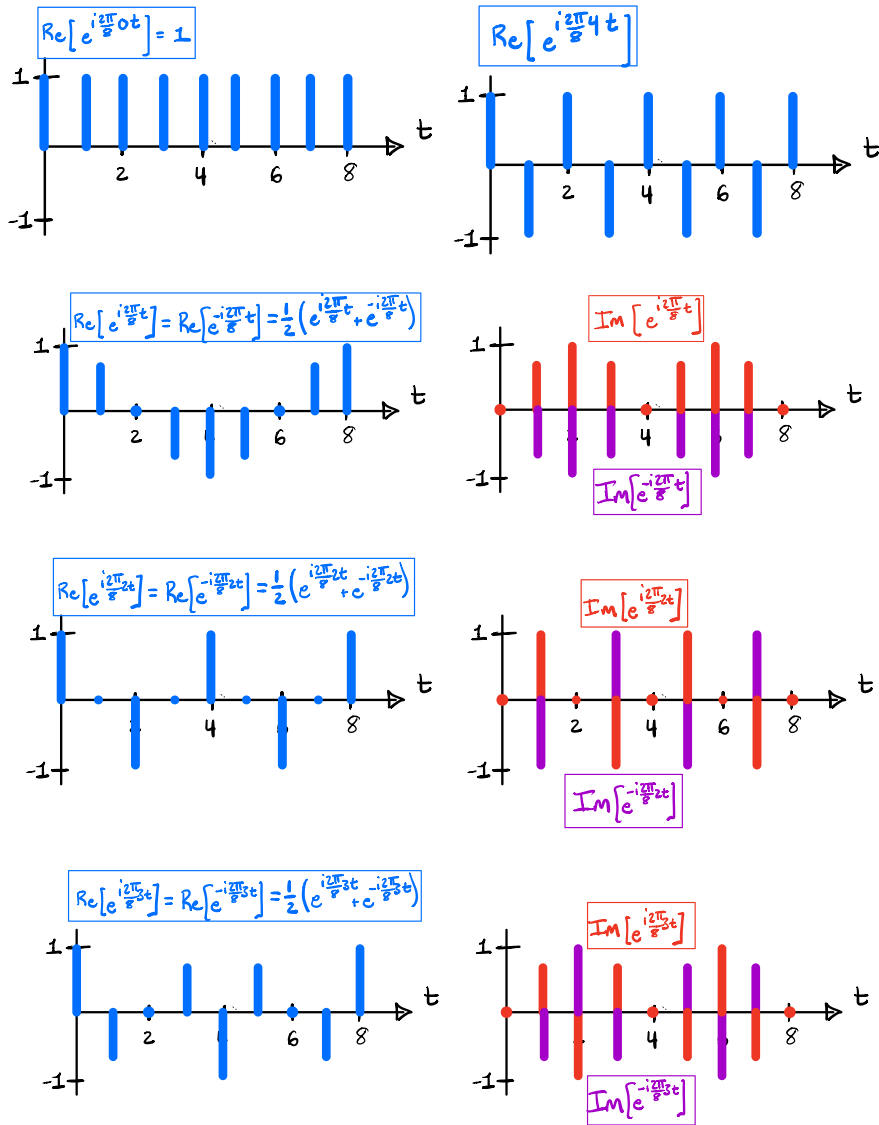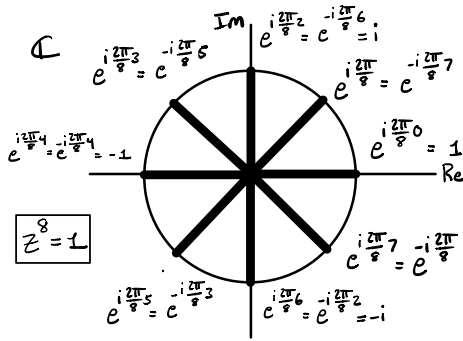   Pairs: $k = (n-1)$ and $-k' = -1$, $k = (n-2)$ and $-k' = -2$, etc.

$\mathbb{C}$ Im $\boxed{z^2 = 1}$

$e^{i\frac{2\pi}{2}} = -1$ π $e^{i\frac{2\pi 0}{2}} = 1$ Re

$e^{i\frac{2\pi}{3}} = e^{-\frac{2\pi}{3}2}$ Im $\boxed{z^3 = 1}$ $e^{i\frac{2\pi 0}{3}} = 1$ Re

$e^{i\frac{2\pi}{3}2} = e^{-i\frac{2\pi}{3}}$

$\boxed{z^4 = 1}$ Im $e^{i\frac{2\pi}{4}} = e^{-i\frac{2\pi}{4}3}$

$e^{i\frac{2\pi}{4}2} = e^{-i\frac{2\pi}{4}2} = -1$ $e^{i\frac{2\pi 0}{4}} = 1$ Re

$e^{i\frac{2\pi}{4}3} = e^{-i\frac{2\pi}{4}}$

$\boxed{z^5 = 1}$

$e^{i\frac{2\pi}{5}2} = e^{-i\frac{2\pi}{5}3}$ Im $e^{i\frac{2\pi}{5}} = e^{-i\frac{2\pi}{5}4}$ $e^{i\frac{2\pi 0}{5}} = 1$ Re

$e^{i\frac{2\pi}{5}3} = e^{-i\frac{2\pi}{5}2}$ $e^{i\frac{2\pi}{5}4} = e^{-i\frac{2\pi}{5}}$

– Roots of unity can be used to define oscillating signals in discrete time.

Let $F^k \in \mathbb{C}^n$ be defined as $\left[F^k\right]_t = e^{\left(i\frac{2\pi k}{n}\right)t}$, ie.

$$F^k = \left[e^{\left(i\frac{2\pi k}{n}\right)0} \quad e^{\left(i\frac{2\pi k}{n}\right)1} \quad \cdots \quad e^{\left(i\frac{2\pi k}{n}\right)(n-1)}\right]^T$$

In discrete time Fourier analysis, we often use the matrix DFT (discrete Fourier transform) matrix $F \in \mathbb{C}^{n \times n}$.

$$F = \begin{bmatrix} F^0 & F^1 & \cdots & F^{n-1} \end{bmatrix}$$

$$= \begin{bmatrix} e^{\left(i\frac{2\pi 0\times 0}{n}\right)} & e^{\left(i\frac{2\pi 0\times 1}{n}\right)} & \cdots & e^{\left(i\frac{2\pi 0\times(n-1)}{n}\right)} \\ e^{\left(i\frac{2\pi 1\times 0}{n}\right)} & e^{\left(i\frac{2\pi 1\times 1}{n}\right)} & \cdots & e^{\left(i\frac{2\pi 1\times(n-1)}{n}\right)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\left(i\frac{2\pi (n-1)\times 0}{n}\right)} & e^{\left(i\frac{2\pi (n-1)\times 1}{n}\right)} & \cdots & e^{\left(i\frac{2\pi (n-1)\times(n-1)}{n}\right)} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\left(i\frac{2\pi 1\times 1}{n}\right)} & \cdots & e^{\left(i\frac{2\pi 1\times(n-1)}{n}\right)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{\left(i\frac{2\pi (n-1)\times 1}{n}\right)} & \cdots & e^{\left(i\frac{2\pi (n-1)\times(n-1)}{n}\right)} \end{bmatrix}$$

24

$$\mathbb{C}$$

$$e^{i\frac{2\pi}{8}3} = e^{-i\frac{2\pi}{8}5}$$
$$e^{i\frac{2\pi}{8}2} = e^{-i\frac{2\pi}{8}6} = i$$
$$e^{i\frac{2\pi}{8}1} = e^{-i\frac{2\pi}{8}7}$$
$$e^{i\frac{2\pi}{8}4} = e^{-i\frac{2\pi}{8}4} = -1$$
$$e^{i\frac{2\pi}{8}0} = 1$$
$$z^8 = 1$$
$$e^{i\frac{2\pi}{8}5} = e^{-i\frac{2\pi}{8}3}$$
$$e^{i\frac{2\pi}{8}6} = e^{-i\frac{2\pi}{8}2} = -i$$
$$e^{i\frac{2\pi}{8}7} = e^{-i\frac{2\pi}{8}}$$

$$Re\left[e^{i\frac{2\pi}{8}0t}\right] = 1$$

$$Re\left[e^{i\frac{2\pi}{8}4t}\right]$$

$$Re\left[e^{i\frac{2\pi}{8}t}\right] = Re\left[e^{-i\frac{2\pi}{8}t}\right] = \tfrac{1}{2}\left(e^{i\frac{2\pi}{8}t} + e^{-i\frac{2\pi}{8}t}\right)$$

$$Im\left[e^{i\frac{2\pi}{8}t}\right]$$
$$Im\left[e^{-i\frac{2\pi}{8}t}\right]$$

$$Re\left[e^{i\frac{2\pi}{8}2t}\right] = Re\left[e^{-i\frac{2\pi}{8}2t}\right] = \tfrac{1}{2}\left(e^{i\frac{2\pi}{8}2t} + e^{-i\frac{2\pi}{8}2t}\right)$$

$$Im\left[e^{i\frac{2\pi}{8}2t}\right]$$
$$Im\left[e^{-i\frac{2\pi}{8}2t}\right]$$

$$Re\left[e^{i\frac{2\pi}{8}3t}\right] = Re\left[e^{-i\frac{2\pi}{8}3t}\right] = \tfrac{1}{2}\left(e^{i\frac{2\pi}{8}3t} + e^{-i\frac{2\pi}{8}3t}\right)$$

$$Im\left[e^{i\frac{2\pi}{8}3t}\right]$$
$$Im\left[e^{-i\frac{2\pi}{8}3t}\right]$$

The columns of F can be used to represent time oscillating signals. In discrete time

Fourier analysis a time (or phase) shift of $\frac{k}{n}$ Hz can be represented by the root of unity $e^{\left(i\frac{2\pi k}{n}\right)}$. Multiplying F^k by $e^{\left(i\frac{2\pi k}{n}\right)}$ shifts each element of the vector up one spot and moves the first element to the end.

$$
e^{\left(i\frac{2\pi k}{n}\right)}F^k = e^{\left(i\frac{2\pi k}{n}\right)}\left[e^{\left(i\frac{2\pi k}{n}\right)0} \quad e^{\left(i\frac{2\pi k}{n}\right)1} \quad \cdots \quad e^{\left(i\frac{2\pi k}{n}\right)(n-1)}\right]^T
$$

$$
= \left[e^{\left(i\frac{2\pi k}{n}\right)1} \quad e^{\left(i\frac{2\pi k}{n}\right)2} \quad \cdots \quad e^{\left(i\frac{2\pi k}{n}\right)(n-1)} \quad e^{\left(i\frac{2\pi k}{n}\right)0}\right]^T
$$

'

Linear combinations and linear dependence

For the following define

$$
x \in \mathbb{R}^n, \qquad y \in \mathbb{R}^m, \qquad A = \begin{bmatrix} | & & | \\ A_1 & \cdots & A_n \\ | & & | \end{bmatrix}
$$

- A vector y is **linear dependent** on the columns of a matrix $A \in \mathbb{R}^{m \times n}$

$$
\text{if } \exists\, x \in \mathbb{R}^n \quad \text{s. t.} \quad y = Ax \qquad \text{or (equivalently) } y = \sum_i A_i x_i
$$

- A set of vectors (the columns of A) is **linearly dependent** if at least one vector is dependent on the others.

$$
A_i = \sum_{j \neq i} A_j x'_j
$$

for some i and $\{x_j\}_{j \neq i}$. A useful characterization is the columns of A are **linearly dependent**

$$
\text{if } \exists\, x \in \mathbb{R}^n, x \neq 0 \quad \text{s. t.} \quad Ax = 0
$$

Indeed for some i

$$
A_i x_i = -\sum_{j \neq i} A_j x_j, \qquad \Rightarrow \qquad A_i = \sum_{j \neq i} A_j \frac{-x_j}{x_i} = \sum_{j \neq i} A_j x'_j
$$

- A vector y is **linearly independent** on the columns of A, if it is not linearly dependent on them, i.e. there does not exist $x \in \mathbb{R}^n$ such that $y = Ax$.

- A set of vectors (cols of A) is **linear independent** if none of the columns are linearly dependent on the others. A useful characterization is

$$Ax = 0 \qquad \Rightarrow \qquad .x = 0$$

- All linear combinations of a set of vectors is the **span** of those vectors.

y, **linearly dependent** on A_1, A_2.

y, **linearly independent** from A_1, A_2.
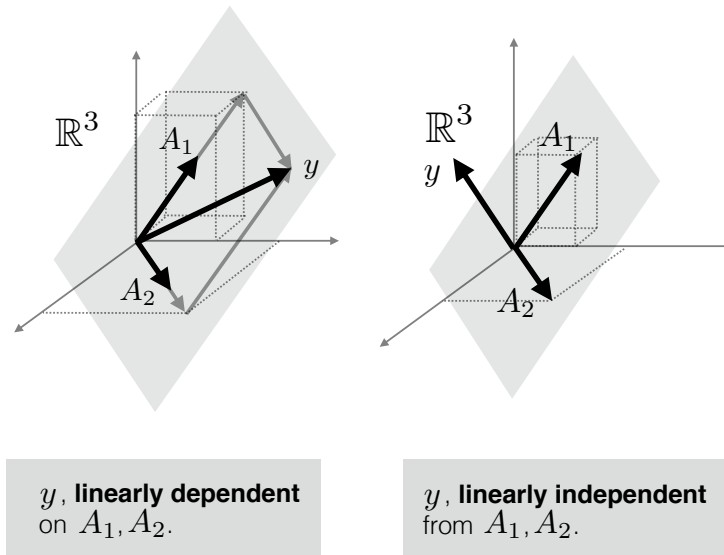
Rank

- **Column rank**: # number of linearly independent columns.

- **Row rank**: # number of linearly independent columns.

- **Rank**: **Column rank=Row rank=Rank**

The column-rank of a matrix is the dimension of the column space of A, $\mathcal{R}(A)$. The row-rank of a matrix is the dimension of the row space, $\mathcal{R}(A^T)$. The column rank and row rank are always equal and are simply called the rank of A, denoted $\mathrm{rk}(A)$.

Column rank = row rank

Proof: Let the column rank be denoted by k and the row rank be denoted by r.

If $A \in \mathbb{R}^{m \times n}$ has column rank k then there exists $C \in \mathbb{R}^{m \times k}$ with linearly independent columns such that

$$A = CV$$

where the columns of $V \in \mathbb{R}^{k \times n}$ are the coordinates of the columns of A with respect to the (basis) columns of C. Thinking of the rows of C as coefficients of linear combinations of the rows of V and realizing that V has k rows, we have that the dimension of $\mathcal{R}(A^T)$ is at most k. Thus we have that $r \leq k$.

If $A \in \mathbb{R}^{m \times n}$ has row rank r then there exists $R \in \mathbb{R}^{r \times n}$ with linearly independent rows such that

$$A = WR$$

where the rows of $W \in \mathbb{R}^{m \times r}$ are the coordinates of the rows of A with respect to the (basis) rows of R. Thinking of the columns or R as coefficients of linear combinations of the columns of W and realizing that W has r columns, we have that the dimension of $\mathcal{R}(A)$ is at most r. Thus we have that $k \leq r$.
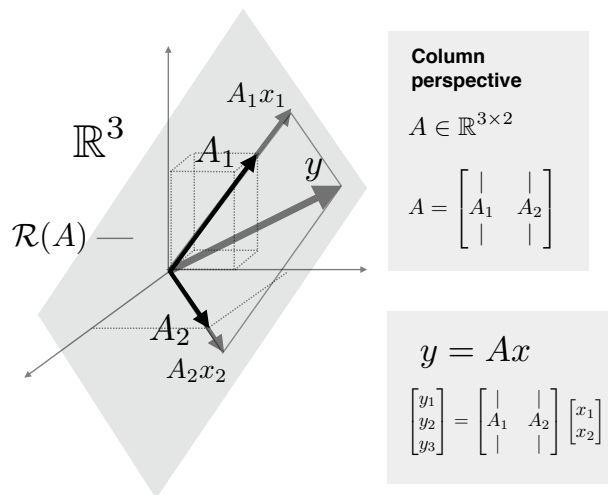
Combining the two inequalities, $r \leq k$ and $k \leq r$ gives that $k = r$, ie. the column and row ranks are equal.
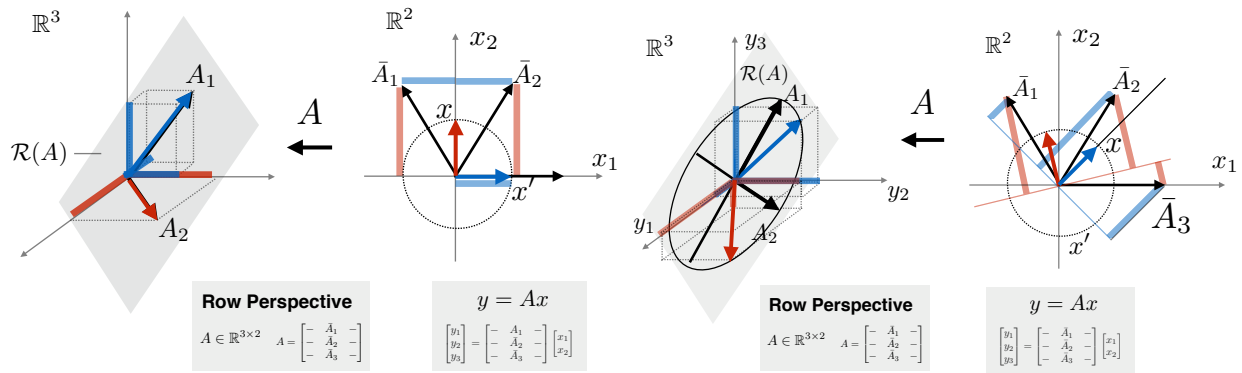
Range and Nullspace

Range Space

A matrix $A \in \mathbb{R}^{m \times n}$ represents a linear map from $\mathbb{R}^n$ which is called the *domain* to $\mathbb{R}^m$ which is called the *co-domain*. The span of the columns of A is a subspace of the co-domain called the *range of A* sometimes denoted $\mathcal{R}(A)$. Note this is equivalent to the definition.

$$\mathcal{R}(A) = \{y \in \mathbb{R}^m \mid y = Ax, \text{ for some } x \in \mathbb{R}^n\} \tag{35}$$

Row Perspective: $A \in \mathbb{R}^{3\times 2}$, $A = \begin{bmatrix} - & \bar{A}_1 & - \\ - & \bar{A}_2 & - \\ - & \bar{A}_3 & - \end{bmatrix}$

$y = Ax$: $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} - & \bar{A}_1 & - \\ - & \bar{A}_2 & - \\ - & \bar{A}_3 & - \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Row Perspective: $A \in \mathbb{R}^{3\times 2}$, $A = \begin{bmatrix} - & \bar{A}_1 & - \\ - & \bar{A}_2 & - \\ - & \bar{A}_3 & - \end{bmatrix}$

$y = Ax$: $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} - & \bar{A}_1 & - \\ - & \bar{A}_2 & - \\ - & \bar{A}_3 & - \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Null Space

The *nullspace of A*, sometimes denoted $\mathcal{N}(A)$, is the subspace of the domain such that

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \tag{36}$$

- **Orthogonal to the rows of A**
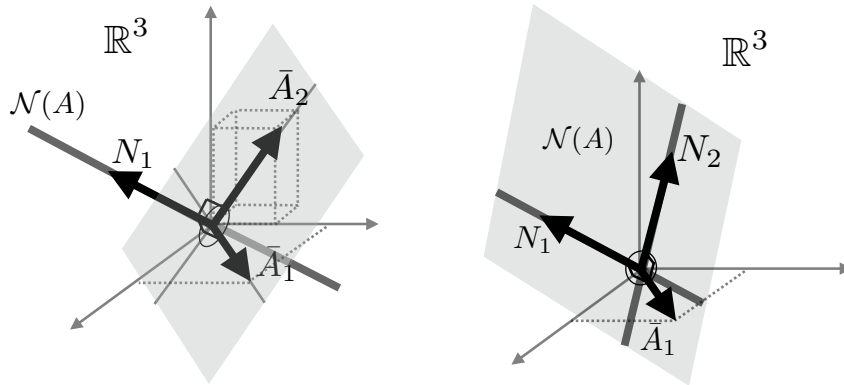
  $\mathcal{N}(A) \perp \mathcal{R}(A^T)$

  $\mathcal{R}(A^T)$ is the span of the rows of A. Decompose $A \in \mathbb{R}^{m\times n}$ into rows as

$$A = \begin{bmatrix} - & \bar{a}_1^T & - \\ & \vdots & \\ - & \bar{a}_1^T & - \end{bmatrix}$$

If $Ax = 0$, then

$$Ax = \begin{bmatrix} - & \bar{a}_1^T & - \\ & \vdots & \\ - & \bar{a}_m^T & - \end{bmatrix} x = \begin{bmatrix} \bar{a}_1^T x \\ \vdots \\ \bar{a}_m^T x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

ie. if $x \in \mathcal{N}(A)$ then x is orthogonal to each row of A. $\mathcal{N}(A) \perp \mathcal{R}(A^T)$ and similarly $\mathcal{N}(A^T) \perp \mathcal{R}(A)$.

29

$$A \in \mathbb{R}^{2 \times 3}$$

$$AN = \begin{bmatrix} - & \bar{A}_1 & - \\ - & \bar{A}_2 & - \end{bmatrix} \begin{bmatrix} | \\ N_1 \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A \in \mathbb{R}^{1 \times 3}$$

$$AN = \begin{bmatrix} - & \bar{A}_1 & - \end{bmatrix} \begin{bmatrix} | & | \\ N_1 & N_2 \\ | & | \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Nullspace: Row perspective

- **Basis Construction:**

  Suppose A has column rank k, ie. $\mathcal{R}(A)$ has dimension k. Assume (without loss of generality) that the first k columns of A are linearly independent (and thus span the range). (A similar construction can be done with any k linearly independent columns of A.) Let

  $$A = \begin{bmatrix} | & & | \\ A_1 & \cdots & A_n \\ | & & | \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} | & & | \\ A_1 & \cdots & A_k \\ | & & | \end{bmatrix}$$

  where $B \in \mathbb{R}^{n \times k}$ is (the first) k linearly independent columns of A. A can then be written as

  $$A = \begin{bmatrix} B & BD \end{bmatrix}$$

  where the columns of $D \in \mathbb{R}^{n \times (n-k)}$ are the coordinates of the remaining $n - k$ columns of A with respect to the columns of B.

  $$\underbrace{\begin{bmatrix} | & & | \\ A_{k+1} & \cdots & A_n \\ | & & | \end{bmatrix}}_{} = BD = \underbrace{\begin{bmatrix} | & & | \\ A_1 & \cdots & A_k \\ | & & | \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} | & & | \\ D_{k+1} & \cdots & D_n \\ | & & | \end{bmatrix}}_{D}$$
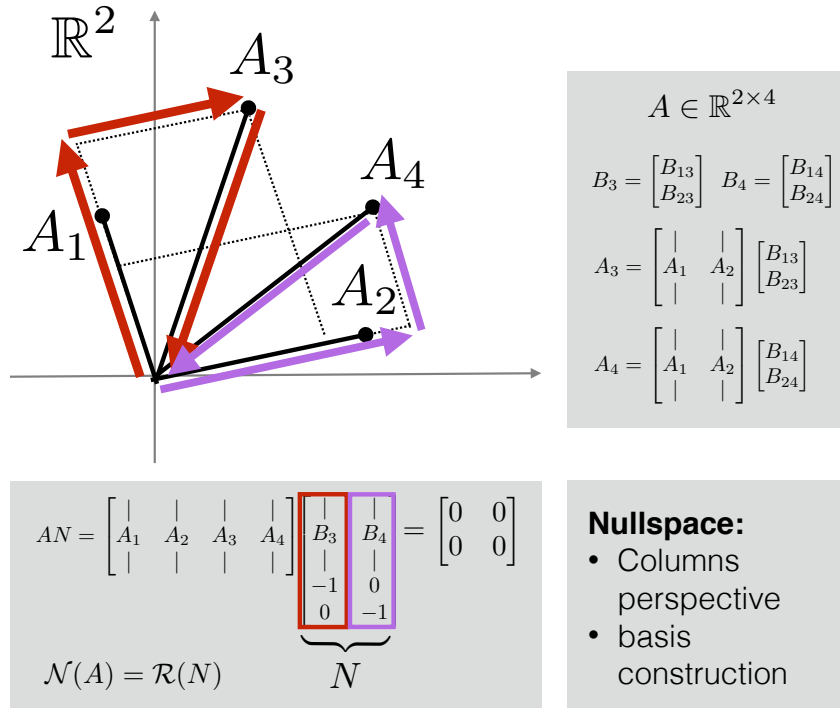
  Let $N \in \mathbb{R}^{n \times (n-k)}$ be given by

  $$N = \begin{bmatrix} -D \\ I \end{bmatrix}$$

30

Note that

$$AN = \begin{bmatrix} B & BD \end{bmatrix} \begin{bmatrix} -D \\ I \end{bmatrix} = 0$$

We have also that the columns of N form a basis for the nullspace of A

$$A \in \mathbb{R}^{2 \times 4}$$

$$B_3 = \begin{bmatrix} B_{13} \\ B_{23} \end{bmatrix} \quad B_4 = \begin{bmatrix} B_{14} \\ B_{24} \end{bmatrix}$$

$$A_3 = \begin{bmatrix} | & | \\ A_1 & A_2 \\ | & | \end{bmatrix} \begin{bmatrix} B_{13} \\ B_{23} \end{bmatrix}$$

$$A_4 = \begin{bmatrix} | & | \\ A_1 & A_2 \\ | & | \end{bmatrix} \begin{bmatrix} B_{14} \\ B_{24} \end{bmatrix}$$

$$AN = \begin{bmatrix} | & | & | & | \\ A_1 & A_2 & A_3 & A_4 \\ | & | & | & | \end{bmatrix} \begin{bmatrix} | & | \\ B_3 & B_4 \\ | & | \\ -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{N}(A) = \mathcal{R}(N) \qquad N$$

Nullspace:
- Columns perspective
- basis construction

Proof:

– **Span $\mathcal{N}(A)$:**
Suppose $Ax = 0$

$$Ax = \begin{bmatrix} B & BD \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$Bx_1 = -BDx_2$$
$$\big((B^T B)^{-1} B^T\big) Bx_1 = -\big((B^T B)^{-1} B^T\big) BDx_2$$
$$x_1 = -Dx_2$$

Note that the linear independence of the columns of B guarantees that $B^T B$ is invertible. Plugging in then gives

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -Dx_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} -D \\ I \end{bmatrix} x_2 = Nx_2$$

showing that x is a linear combination of the columns of N.

- **Linear independence:**
  Suppose $Nx_2 = 0$

$$Nx_2 = \begin{bmatrix} -D \\ I \end{bmatrix} x_2 = \begin{bmatrix} -Dx_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

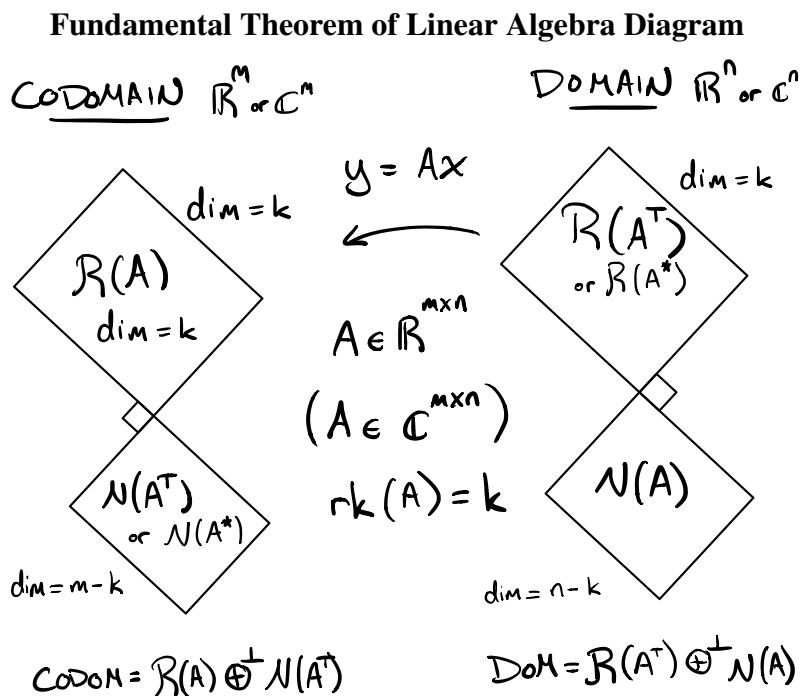It follows that $x_2 = 0$.

Rank-Nullity Theorem

The explicit construction of a basis for the nullspace given above shows that if a matrix has (column) rank k then the nullspace has dimension $n - k$. The dimension of the nullspace is known as the nullity and we have the rank-nullity theorem
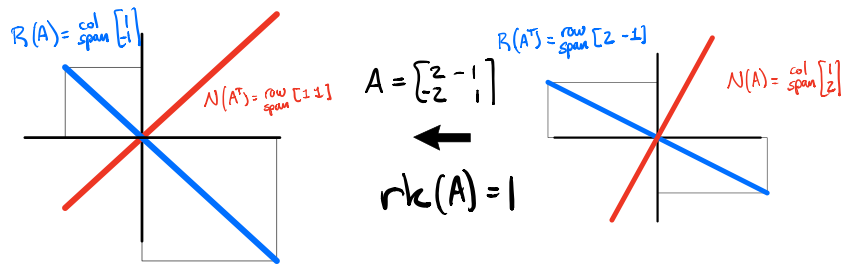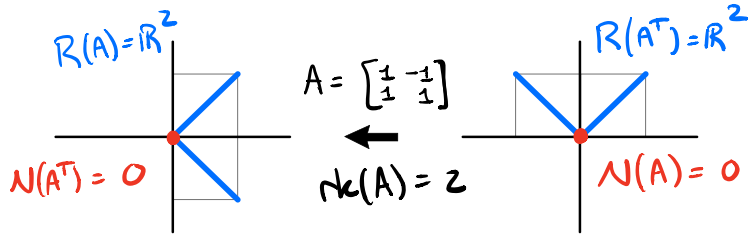
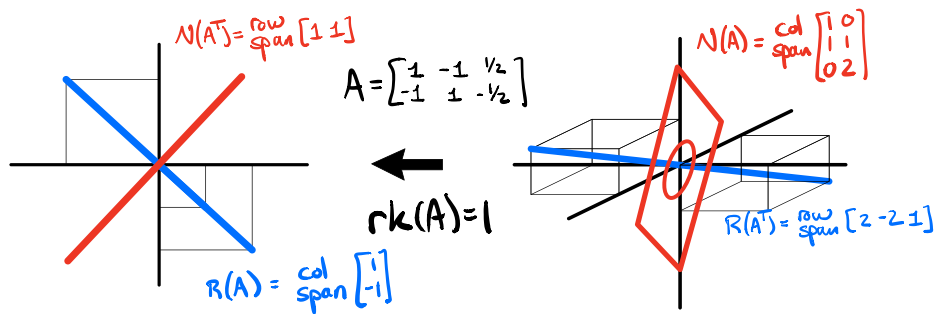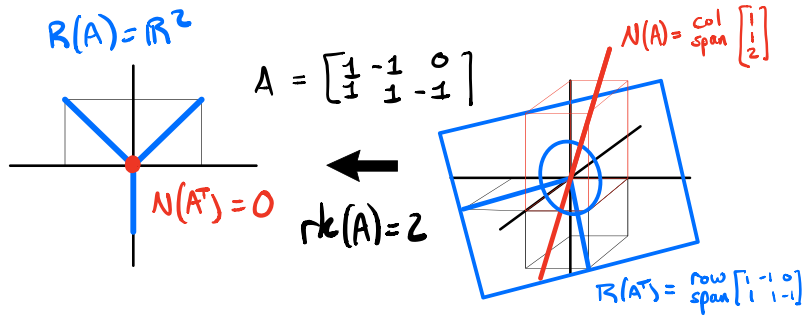$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n$$
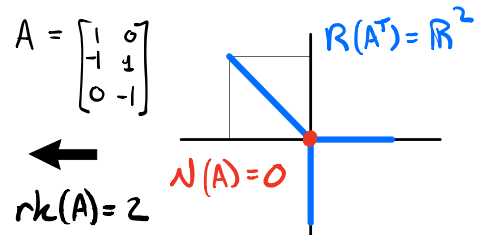$$\mathrm{rk}(A) + \dim(\mathcal{N}(A)) = n$$

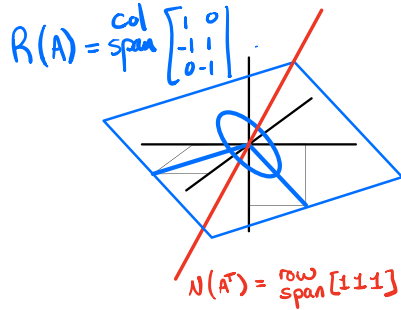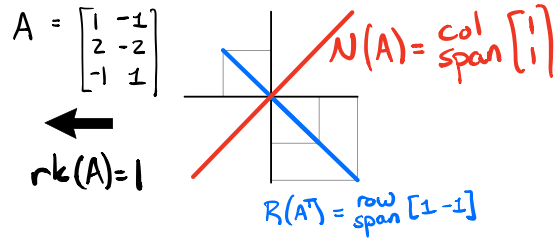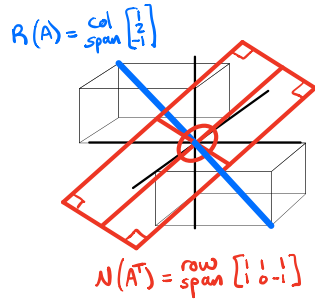# 6 Fundamental Theorem of Linear Algebra

$\mathcal{R}(A^T)$ and $\mathcal{N}(A)$ are *orthogonal subspaces* of the domain, meaning that any vector in one is orthogonal to any vector in the other. In addition, together $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$ span all of the domain $\mathbb{R}^n$. Similarly, $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are orthogonal subspaces of the co-domain and together they span the co-domain.
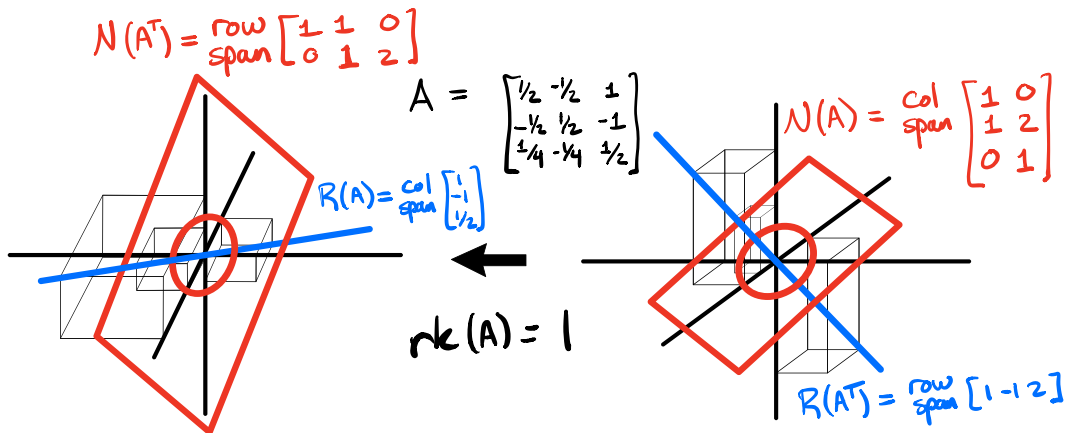
Fundamental Theorem of Linear Algebra Diagram

CODOMAIN $\mathbb{R}^m$ or $\mathbb{C}^m$ DOMAIN $\mathbb{R}^n$ or $\mathbb{C}^n$

$y = Ax$

$\dim = k$ $\dim = k$

$\mathcal{R}(A)$ $\mathcal{R}(A^T)$ or $\mathcal{R}(A^*)$

$\dim = k$ $A \in \mathbb{R}^{m \times n}$

$\left(A \in \mathbb{C}^{m \times n} \right)$

$\mathcal{N}(A^T)$ or $\mathcal{N}(A^*)$ $\mathrm{rk}(A) = k$ $\mathcal{N}(A)$

$\dim = m - k$ $\dim = n - k$

$\mathrm{CODOM} = \mathcal{R}(A) \oplus^\perp \mathcal{N}(A^T)$ $\mathrm{DOM} = \mathcal{R}(A^T) \oplus^\perp \mathcal{N}(A)$

32

$R(A) = \mathbb{R}^2$

$N(A^T) = 0$

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

$rk(A) = 2$

$R(A^T) = \mathbb{R}^2$

$N(A) = 0$

$R(A) = \begin{smallmatrix} col \\ span \end{smallmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$N(A^T) = \begin{smallmatrix} row \\ span \end{smallmatrix} [1 \; 1]$

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

$rk(A) = 1$

$R(A^T) = \begin{smallmatrix} row \\ span \end{smallmatrix} [2 \; -1]$

$N(A) = \begin{smallmatrix} col \\ span \end{smallmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

2×2 **examples of rank 2 and 1.**

$R(A) = \mathbb{R}^2$

$N(A^T) = 0$

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

$rk(A) = 2$

$N(A) = \begin{smallmatrix} col \\ span \end{smallmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$

$R(A^T) = \begin{smallmatrix} row \\ span \end{smallmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \end{bmatrix}$

$N(A^T) = \begin{smallmatrix} row \\ span \end{smallmatrix} [1 \; 1]$

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

$rk(A) = 1$

$R(A) = \begin{smallmatrix} col \\ span \end{smallmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$N(A) = \begin{smallmatrix} col \\ span \end{smallmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix}$

$R(A^T) = \begin{smallmatrix} row \\ span \end{smallmatrix} [2 \; -2 \; 1]$

$R(A) = \underset{\text{span}}{\text{col}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$

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

$N(A) = \underset{\text{span}}{\text{col}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$rk(A) = 1$

$N(A^T) = \underset{\text{span}}{\text{row}} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix}$

$R(A^T) = \underset{\text{span}}{\text{row}} \begin{bmatrix} 1 & -1 \end{bmatrix}$

$R(A) = \underset{\text{span}}{\text{col}} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$

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

$R(A^T) = \mathbb{R}^2$

$rk(A) = 2$

$N(A) = 0$

$N(A^T) = \underset{\text{span}}{\text{row}} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$

2×3 **examples of rank 2 and 1.** 3×2 **examples of rank 2 and 1.**

34

$R(A) = \mathbb{R}^3$

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

$R(A^T) = \mathbb{R}^3$

$N(A^T) = 0$

$N(A) = 0$

$rk(A) = 3$

$N(A^T) = \underset{span}{row} [1\ 1\ 1]$

$N(A) = \underset{span}{col} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

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

$rk(A) = 2$

$R(A) = \underset{span}{col} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$

$R(A^T) = \underset{span}{row} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$

$N(A^T) = \underset{span}{row} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

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

$N(A) = \underset{span}{col} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$

$R(A) = \underset{span}{col} \begin{bmatrix} 1 \\ -1 \\ 1/2 \end{bmatrix}$

$rk(A) = 1$

$R(A^T) = \underset{span}{row} [1\ -1\ 2]$

3×3 examples of rank 3, 2, and 1.

Systems of Equations

Matrices are used to represent and solve systems of linear equations. Suppose we $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ that satisfy.

$$y = Ax \tag{37}$$

Note that this equation is slightly more complicated than it first appears. Depending on the shape of A it may have a unique solution, no solution, or a whole subspace of solutions.

Unique Solution

The simplest case is that A is square, ie. $x, y \in \mathbb{R}^n$ and the columns are linearly independent. This means there is a unique linear commbination of the columns that reaches every individual point y in the co-domain. We can compute this exact linear combination by doing *Gaussian elimination* also known as *row reduction*. Each step of Gaussian elimination, each *elementary row operation* can be represented by left-multiplication of Equation (37) by a specific type of matrix called *elementary matrices*. These elementary matrices come in three types: row-multiplying, row-swapping, and row-adding demonstrated below

$$\underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \alpha & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{multiplying a row by } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & \alpha & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{adding a row times } \alpha}, \quad \underbrace{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}_{\text{swapping rows}}$$

$$\tag{38}$$

When we perform Gaussian elimination on Equation (37) to transform A into the identity, we left-multiply by the appropriate set of elementary matrices $\{E_1, \ldots, E_k\}$

$$\underbrace{(E_k \cdots E_1)}_{A_l^{-1}} y = \underbrace{(E_k \cdots E_1)A}_{I} x \tag{39}$$

These elementary matrices multiplied together are called the *left-inverse* $A_l^{-1} = (E_k \cdots E_1)$, ie. the matrix that transforms A into the identity by left-multiplying. Note that we could have performed a similar procedure to solve the equation $y^\top = x^\top A$ except we would multiply on the right by *elementary column matrices*. This procedure would construct the *right inverse* of A, denoted A_r^{-1}. $y^\top A_r^{-1} = x^\top A A_r^{-1} = x^\top$. Assuming A is square and invertible, these two left and right inverses

are the same and we simply denote them as $A^{-1} = A_l^{-1} = A_r^{-1}$. This can be seen from

$$A_l^{-1} \cdot A = I$$
$$A_l^{-1} \cdot A \cdot A_r^{-1} = I \cdot A_r^{-1} \qquad (40)$$
$$A_l^{-1} = A_r^{-1}$$

No solution (Least Squares)

If $m > n$, ie. A is "tall", then it is unlikely that there is any solution at all. The columns of A span a subspace of the co-domain called the range of A. There will only be a solution for x if y happens to lie in this subspace. If the columns of A are linearly independent, then A will still have a left-inverse. This is based on the fact that the linear independence of the columns of implies that the matrix $A^\top A$ will be invertible. This in turn implies that we can construct a left-inverse as $A_l^{-1} = (A^\top A)^{-1} A^\top$. Supposing that y is actually in the range of A, ie. there does exist an x solving (37), we can find this x using this left-inverse.

$$\text{Assume } y \text{ in range of } A... \qquad y = Ax$$
$$(A^\top A)^{-1} A^\top y = (A^\top A)^{-1} A^\top \cdot Ax = x \qquad (41)$$

Now suppose y is not in the range of A. We can still try to find an x that makes Ax as close to y as possible, ie. we can try to minimize

$$\|y - Ax\|_2^2 = (y - Ax)^\top (y - Ax) = y^\top y + y^\top Ax + x^\top A^\top Ax = \sum_i (y_i - A_{i:}x)^2 \qquad (42)$$

x that minimizes this quantity is called the *least squares solution*, $x_{\text{lsq}} = A(A^\top A)^{-1} A^\top y$ which is the projection of y onto the range of A. We can derive the least squares solution by computing the derivative of (42) and set it equal to 0.

$$\frac{\partial}{\partial x}\left(y^\top y - y^\top Ax - x^\top Ay + x^\top A^\top Ax\right) = -2y^\top A + 2x^\top A^\top A = 0 \qquad (43)$$
$$\Rightarrow \qquad x = (A^\top A)^{-1} A^\top y \qquad (44)$$

No x s.t.

$y = Ax$

$A = [\vec{A_1}, \vec{A_2}]$

Nullspace of $A^\top$

$y - Ax$

Range of A

$x_{\text{lsq}} = A(A^\top A)^{-1} A^\top x$

minimizes $|y - Ax|^2$

37

Subspace/Continuum of Solutions

Suppose $n > m$, ie. A is "fat", and there are more than m linearly independent columns. In . this case, we have more columns than we need to span the space. If we pick any m linear independent columns, we can compute a solution. Suppose the first m columns of A are linearly independent, $A = [\bar{A} \cdots]$ where $\bar{A} \in \mathbb{R}^{m \times m}$. We can then compute one solution as $x^1 = [\bar{A}^{-1}y \; \mathbf{0}]^\top$ where $\mathbf{0}$ is the appropriate size vector of zeros. The same procedure with different sets of columns produces up to $n - m + 1$ linearly independent solutions which we can organize as the columns of $X = [x^1 \cdots x^{n-m+1}]$. Note that $A(x^i - x^j) = 0$, ie. $x^i - x^j$ is in the nullspace of A. A basis for the nullspace of A can be computed as the columns of XW where the matrix $W \in \mathbb{R}^{(n-m+1) \times (n-m)}$ is given by $W = [\mathbf{1} - I]^\top$ where $\mathbf{1}$ is a vector of ones of the appropriate size. (Note that W computes differences between the columns of X. A different W that computes column differences could be used.) Any solution of (37) has the form

$$x = x^0 + x_{\mathrm{NS}} = x^0 + XWz$$

for some $z \in \mathbb{R}^{n-m}$, ie. any solution consists of some specific solution x^0 plus some component in the nullspace of A. We can compute a specific solution using the method above (selecting m linearly independent columns). However, assuming the rows of A are linearly independent and if we want a specific solution x^0 that is orthogonal to the nullspace of A, then we can select x as a linear combination of the rows of A. Assume x^0 has the form $x^0 = A^\top w$ with $w \in \mathbb{R}^m$. Plugging into (37), gives

$$y = AA^\top w \qquad \Rightarrow \qquad w = (AA^\top)^{-1}y \qquad \Rightarrow \qquad x^0 = A^\top(AA^\top)^{-1}y \qquad (45)$$

Note that x^0 is y times a right-inverse of A. Note also that x^0 is orthogonal to the nullspace of A since $x_{NS}^\top A^\top(AA^\top)^{-1} = 0$. Note also that x^0 computed in this way is the solution with the *minimum 2-norm*. To see this, note that adding some component from the nullspace only increases the square of the 2-norm.

$$|x^0 + x_{\mathrm{NS}}|^2 = (x^0 + x_{\mathrm{NS}})^\top (x^0 + x_{\mathrm{NS}}) \tag{46}$$
$$= (x^0)^\top x^0 + 2x_{\mathrm{NS}}^\top x^0 + x_{\mathrm{NS}}^\top x_{\mathrm{NS}} \tag{47}$$
$$= (x^0)^\top x^0 + x_{\mathrm{NS}}^\top x_{\mathrm{NS}} = |x^0|^2 + |x_{\mathrm{NS}}|^2 \geq= |x^0|^2 \tag{48}$$

General Case

Minimum-Norm, Least Squares (Moore-Penrose Pseudoinverse)

In the general case, $A \in \mathbb{R}^{m \times n}$ may not be full column or row rank. In this case neither $A^\top A$ or $AA^\top$ or In this case, for there are many possible x's that are all equally bad at reaching y. Perhaps the most sensible x to choose in this case is the *minimum-norm, least squares* solution. Here, we look for the least squares solution that does not include any element in the nullspace of A. This can be computed using the *Moore-Penrose* pseudoinverse denoted $A^\dagger$. This is best understood using

the singular-value decomposition. (Here we assume the matrix A is real and so we use the real SVD; an exactly analogous formula works in the complex case). Given that the SVD of A, $A^\dagger$ can be written as follows

$$A = \underbrace{\begin{bmatrix} U_1\ U_2 \end{bmatrix}}_{U} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -V_1^\top- \\ -V_2^\top- \end{bmatrix}}_{V^\top} \qquad \Longrightarrow \qquad A^\dagger = \underbrace{\begin{bmatrix} V_1\ V_2 \end{bmatrix}}_{V} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -U_1^\top- \\ -U_2^\top- \end{bmatrix}}_{U^\top} \qquad (49)$$

Note that in this formula, we've followed the standard rules for taking an inverse—reversing the order and inverting V and U (since $U^{-1} = U^\top$ and $V^{-1} = V^\top$) — but we only inverted the part of the center matrix that is invertible. As detailed in the SVD lecture, U_1, U_2, V_1, V_2 have the following interpretations.

$$U_1 : \text{orthonormal basis for the range of } A$$
$$U_2 : \text{orthonormal basis for the nullspace of } A^\top$$
$$V_1 : \text{orthonormal basis for the range of } A^\top$$
$$V_2 : \text{orthonormal basis for the nullspace of } A$$

Intuitively, A contains an invertible map between the range of A and the range of $A^\top$ and we've inverted this part of A while ignoring the part in the nullspace. Expanding out, we get that (49) could be written as

$$A = U_1 \Sigma V_1^\top \qquad \Longrightarrow \qquad A^\dagger = V_1 \Sigma^{-1} U_1^\top \qquad (50)$$

Note that here this looks like a simple formula except U_1 and V_1 are tall so they can't simply be inverted.

Exercise:. Show that for an equation $y = Ax$ (for general A), $x = A^\dagger y$ gives the least squares solution with the minimum norm.

Gaussian Elimination: Row Reduction

We now consider what happens if we perform Gaussian elimination on a general matrix with rank k where $k < m$, $k < n$. For a matrix $A \in \mathbb{R}^{m \times n}$ with rank k, assuming the first k columns of A are linearly independent we can find an invertible $E \in \mathbb{R}^{m \times m}$ that is a composition of elementary matrices $E = E_\ell \cdots E_1$ such that

$$EA = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \qquad (51)$$

with $I \in \mathbb{R}^{k \times k}$ and $B \in \mathbb{R}^{k \times n-k}$. It will be helpful to decompose E and also E^{-1} as

$$E = \begin{bmatrix} -\ E' \ - \\ -\ E'' \ - \end{bmatrix}, \qquad E^{-1} = \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix}$$

39

where

$$E' \in \mathbb{R}^{k \times m}, \quad E'' \in \mathbb{R}^{(m-k) \times m}, \qquad F' \in \mathbb{R}^{m \times k}, \quad F'' \in \mathbb{R}^{m \times (m-k)}$$

Note: since A has rank k, there will always be at least k linearly independent columns, if the first k columns aren't linearly independent then the above formula must be changed to be

$$EAP = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \tag{52}$$

where P is some permutation matrix that reorders the columns so that the first k are linearly independent. This is the most general form Gaussian elimination can take. In this case, we solve the linear system $y = APx'$ where $x = Px' \iff x' = P^\top x$. Once we've solved for x', we can recover x. For simplicity, we will consider equation (51).

Note that since the columns (or rows) of E', E'', F', F'' are all columns (or rows) of invertible matrices, they must be linearly independent. We note also that

$$I = EE^{-1} = \begin{bmatrix} - & E' & - \\ - & E'' & - \end{bmatrix} \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix} = \begin{bmatrix} E'F' & E'F'' \\ E''F' & E''F'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Specifically, note which submatrices must be orthogonal. We also have that $I = E^{-1}E = F'E' + F''E''$.

Using the above decomposition the row-reduction operations become

$$EA = \begin{bmatrix} - & E' & - \\ - & E'' & - \end{bmatrix} \begin{bmatrix} | & | \\ A' & A'' \\ | & | \end{bmatrix} = \begin{bmatrix} E'A' & E'A'' \\ E''A' & E''A'' \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix}$$

It can also be quite useful to write the above equation as a decomposition of A

$$A = E^{-1} \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} | & | \\ F' & F'' \\ | & | \end{bmatrix} \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} = F' \begin{bmatrix} I & B \end{bmatrix}$$

It is clear from this that F' must span the range of A. Since the columns are linearly independent it is also a basis. We also have that the rows of E'' form a basis for the nullspace of A^T by a similar linear independence argument, rank-nullity (applied to the co-domain) and the fact that $E''F' = 0$. Note that the rows of $\begin{bmatrix} I & B \end{bmatrix}$ are also linearly independent (since the first subblock is the identity) and thus $\begin{bmatrix} I & B \end{bmatrix}^T$ is a basis for the range of A^T Finally, by arguments given in the discussion on nullspaces, the rows of $\begin{bmatrix} B^T & -I \end{bmatrix}^T$ are a basis for the nullspace of A. We can summarize these insights in a list of bases for the four fundamental subspaces related to A.

$$\underset{of\ A}{Range} : \ F', \qquad \underset{of\ A^T}{Nullspace} : \ E''^T \qquad \underset{of\ A^T}{Range} : \ \begin{bmatrix} I \\ B^T \end{bmatrix}, \qquad \underset{of\ A}{Nullspace} : \ \begin{bmatrix} B \\ -I \end{bmatrix}$$

40

Gaussian Elimination: Column Reduction

We can make a similar argument to the above for column reduction in the general case. For a matrix $A \in \mathbb{R}^{m \times n}$ with rank k assuming the first k rows of A are linearly independent, we can find an invertible $E \in \mathbb{R}^{m \times m}$ that is a composition of elementary matrices $E = E_1 \cdots E_\ell$ such that

$$AE = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix}$$

with $I \in \mathbb{R}^{k \times k}$ and $C \in \mathbb{R}^{k \times m-k}$. (If the first k rows are not linearly independent, replace AE with PAE for a permutation matrix P that shuffles k linearly independent rows into the first k spots.) Note here this composition of elementary matrices E will be different than in the row reduction case. Again it will be helpful to decompose E and also E^{-1} as

$$E = \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix}, \qquad E^{-1} \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix}$$

where $E' \in \mathbb{R}^{n \times k}$, $E'' \in \bowtie \times \bowtie - \urcorner$, $F' \in \mathbb{R}^{k \times n}$, and $F'' \in \bowtie - \urcorner \times \bowtie$. Note that since the columns (or rows) of E', E'', F', F'' are all columns (or rows) of invertible matrices, they must be linearly independent. We note also that

$$I = E^{-1}E = \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix} \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix} = \begin{bmatrix} F'E' & F'E'' \\ F''E' & F''E'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

Specifically, note which submatrices must be orthogonal. We also have that $I = EE^{-1} = E'F' + E''F''$.

Using the above decomposition the column-reduction operations become

$$AE = \begin{bmatrix} - & A' & - \\ - & A'' & - \end{bmatrix} \begin{bmatrix} | & | \\ E' & E'' \\ | & | \end{bmatrix} = \begin{bmatrix} A'E' & A'E'' \\ A''E' & A''E'' \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix}$$

It can also be quite useful to write the above equation as decomposition of A

$$A = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix} E^{-1} = \begin{bmatrix} I & 0 \\ C & 0 \end{bmatrix} \begin{bmatrix} - & F' & - \\ - & F'' & - \end{bmatrix} = \begin{bmatrix} I \\ C \end{bmatrix} F'$$

It is clear from this that the rows of F' must span the range of A^T. Since the rows are linearly independent it is also a basis. We also have that the columns of E'' form a basis for the nullspace of A by a similar linear independence argument, rank-nullity (applied to the co-domain) and the fact that $F'E'' = 0$. Note that the columns of $\begin{bmatrix} I \\ C \end{bmatrix}$ are also linearly independent (since the first subblock is the identity) and thus $\begin{bmatrix} I \\ C \end{bmatrix}^T$ is a basis for the range of A Finally, by arguments given

41

in the discussion on nullspaces, the rows of $\begin{bmatrix} C & -I \end{bmatrix}$ are a basis for the nullspace of A^T. We can summarize these insights in a list of bases for the four fundamental subspaces related to A.

$$\begin{matrix} Range \\ of\ A \end{matrix} : \begin{bmatrix} I \\ C \end{bmatrix}, \quad \begin{matrix} Nullspace \\ of\ A^T \end{matrix} : \begin{bmatrix} C^T \\ -I \end{bmatrix} \quad \begin{matrix} Range \\ of\ A^T \end{matrix} : F'^T, \quad \begin{matrix} Nullspace \\ of\ A \end{matrix} : E''$$

Inverse Properties

Properties of inverses:

$P, Q \in \mathbb{C}^{n \times n}$ invertible, and $k \in \mathbb{C}$.

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

- $(kP)^{-1} = \frac{1}{k} P^{-1}$

- $(PQ)^{-1} = Q^{-1} P^{-1}$

- $\det(P^{-1}) = \frac{1}{\det(P)}$

- $P^{-1} = \frac{1}{\det(P)} \text{Adj}(P)$

Equivalent Inverse Properties:

- P is invertible, ie. P^{-1} exists.

- $P^\top$ is invertible

- P can be row reduced to the identity (via Gaussian Elimination (GE))

- P can be column reduced to the identity (via GE).

- P is a product of elementary matrices.

- P (square) is full row rank.

- P (square) is full column rank.

- Columns of P (square) are linearly independent, ie. $Px = 0 \implies x = 0$.

- Rows of P (square) are linearly independent, ie. $y^\top P = 0 \implies y^\top = 0$. Rows of P (square) are linearly independent.

- $y = Px$ has a unique solution for each y.

- P has a trivial nullspace. $\mathcal{N}(P) = \{0\}$

- 0 is not an eigenvalue of P.

- $\det(P) \neq 0$.

- There exists Q such that $PQ = QP = I$ ($P^{-1} = Q$).

- P has a left and a right inverse.

Inverse Formulas

- 2×2 inverse

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad P^{-1} = \frac{1}{\det(P)}\text{Adj}(P) = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{\det(P)}\Big[\text{Tr}(P)I - P\Big]$$

- 3×3 inverse

$$P^{-1} = \frac{1}{\det(P)}\text{Adj}(P)$$

$$= \frac{1}{\det(P)}\left[\tfrac{1}{2}\Big(\text{Tr}(P)^2 - \text{Tr}(P^2)\Big)I - P\text{Tr}(P) + P^2\right]$$

- **Block Matrix Inversion**

$$\mathbf{P}^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} assuming \mathbf{D}^{-1}$ and $(A-$

  $BD^{-1}C)^{-1}$ exist or A^{-1} and $(D - CA^{-1}B)^{-1}$ exist.

  Proof:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left(\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}\right)^{-1}$$

$$= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left(\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}\right)^{-1}$$

$$= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

43

- **Woodbury Matrix Identity**
  Note: this formula is a work horse of matrix algebra and worth memorizing.

  $$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

  where $A \in \mathbb{C}^{n \times n}$, $U \in \mathbb{C}^{n \times k}$, $C \in \mathbb{C}^{k \times k}$, and $V \in \mathbb{C}^{k \times n}$. This formula is particularly useful when $n > k$ (U is tall and V is fat). In particular, if U is a column vector, V is a row vector, and C is a scalar, then this equation is called the *Sherman-Morrison Formula*.

  Special Cases:

  – **Inverse of $A + B$:**

  $$(A + B)^{-1} = A^{-1} - A^{-1}B(I + A^{-1}B)^{-1}A^{-1}$$

  Note: other forms are possible as well.

  – **Sherman-Morrison:**

  $$(A + uv^\top)^{-1} = A^{-1} - A^{-1}u\frac{1}{1 + v^\top A^{-1}u}v^\top A^{-1}$$

- **Neumann Series**

  $$A^{-1} = \sum_{n=0}^{\infty}(I - A)^n, \qquad \text{if } \lim_{n \to \infty}(I - A)^n = 0$$

- **Derivative of Inverse**
  For $P(t)$

  $$\frac{\partial P^{-1}}{\partial t} = -P^{-1}\frac{\partial P}{\partial t}P^{-1}$$

Basis

A set of vectors $\{V_i\}_{i=1}^n$ are a **basis** for a vector space $\mathcal{V}$ if 1.) $\{V_i\}_{i=1}^n$ span $\mathcal{V}$ and 2) $\{V_i\}_{i=1}^n$ are linearly independent. *Every basis for $\mathcal{V}$ has the same number of vectors and the number of vectors in a basis for $\mathcal{V}$ is called the **dimension** of $\mathcal{V}$.*

Change of Basis

Suppose we have a basis for $\mathbb{R}^n$ stored in the columns of a square matrix $P \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$. The *coordinates* of x with respect to the basis P are the coefficients

of the linear combination of the columns of P that gives the vector x. We represent these coordinates in a vector $z \in \mathbb{R}^n$ and have that $x = Pz$. Note that x is really the coordinates of itself with respect to the *standard basis vectors* which are the columns of the identity matrix, ie. $x = Ix$.

To compute the coordinates of x with respect to P, we simply invert P, ie. $z = P^{-1}x$. If the columns of P are a basis, then they must be linearly independent and P is invertible.

For simple 2×2 and 3×3 cases, we could also compute the coordinates z by drawing the basis vectors and x and eyeballing the appropriate coordinates as in the figure above. This is a useful exercise in order to develop intuition coordinates. We give several examples in the figures below. (Once you guess z, you can always easily check how close you are by computing Pz and comparing it with x.)

$$x = \begin{bmatrix} | & | \\ P_1 & P_2 \\ | & | \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad x = \begin{bmatrix} | & | \\ P_1 & P_2 \\ | & | \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad x = \begin{bmatrix} | & | \\ P_1 & P_2 \\ | & | \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

Similarity Tranforms

Now suppose, we have a square matrix $A \in \mathbb{R}^{n \times n}$ that transforms the vector $x \in \mathbb{R}^n$ into a vector $y \in \mathbb{R}^n$, ie. $y = Ax$. We represent both x and y in terms of a new basis given by the columns of $P \in \mathbb{R}^{n \times n}$, ie. $x = Px'$ and $y = Py'$. We want to find a matrix A' that represents the action of A when x and y are expressed in the x' and y' coordinates respectively, ie. we want to find $A' \in \mathbb{R}^{n \times n}$ such that $y' = A'x'$. We can do this by plugging in the relationships between x and z

$$y = Ax \tag{53}$$
$$Py' = APx' \tag{54}$$
$$y' = \underbrace{P^{-1}AP}_{A'} x' \tag{55}$$

We say that A' is related to A by a *similarity transform*, ie. A' represents the transformation of A just with respect to a different coordinate system. The construction of A' is illustrated in the figure below.

$$A' = P^{-1} \, A \, P$$

Similarity Transform: $A \in \mathbb{R}^{n \times n}$

A generalization of this concept is that $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are transformed under different coordinate transformations, ie. $x = Px'$ and $y = Qy'$. A is then transformed as

$$y = Ax \tag{56}$$
$$Qy' = APx' \tag{57}$$
$$y' = \underbrace{Q^{-1}AP}_{A'} x' \tag{58}$$

Note that here, it is not necessary that A be square and x and y have the same dimension. This situation is illustrated in the figure below.

$$A' = Q^{\text{-}1}\, A\, P$$

Coordinate Transform: $A \in \mathbb{R}^{m \times n}$

Traces and Determinants

Two useful numbers associated with square matrices are the *trace* and the *determinant*. The trace is the sum of the diagonals

$$\text{Tr}(A) = \sum_i A_{ii} \tag{59}$$

Traces are very well behaved algebraic. One can check immediately the following identities.

$$\text{Tr}(A) = \text{Tr}(A^T), \qquad \text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B), \qquad \text{Tr}(AB) = \text{Tr}(BA) \tag{60}$$

Formulas for the determinant are generally complicated but they compute how the volume of the unit cube changes under the transformation A.

$$\det(A) = \text{signed volume of the unit cube transformed by } A \tag{61}$$

The sign of the determinant flips if the unit cube is reflected across some axis.

Determinants have the properties

$$\det(A) = \det(A^T), \qquad \det(A^{-1}) = \det(A)^{-1}, \qquad \det(AB) = \det(BA) = \det(A)\det(B) \tag{62}$$

Both the trace and determinant have special relationships with the eigenvalues of A (see below for discussion of eigenvalues). If the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$ then we have that

$$\mathrm{Tr}(A) = \sum_i \lambda_i, \qquad \det(A) = \prod_i \lambda_i \tag{63}$$

Eigenvectors, Eigenvalues, and Diagonalization

In general, multiplying a column vector $x \in \mathbb{R}^n$ by a square matrix $A \in \mathbb{R}^{n \times n}$ causes that vector to stretch and to rotate. However, some vectors in specific subspaces are *only stretched, not rotated*. Another way to say this is that those subspaces are *invariant* with respect to A. These invariant subspaces are called *right eigenspaces* and vectors within them are called *right eigenvectors*. The amount each eigenvector is stretched is called it's *eigenvalue*. We can also consider a similar situation where left multiplying A by specific row vectors only causes them to stretch. These row vectors are called *left eigenvectors* and they live in *left eigenspaces*. (The eigenvalues for left and right eigenvectors turn out to be the same, ie. left and right eigenspaces come in pairs.) Finding a linearly independent sets of eigenvectors (either left or right) for a square matrix A is one of the fundamental problems of linear algebra. **If we represent vectors as coordinates with respect to a basis of eigenvectors, then the action of the matrix simply becomes scaling each individual coordinate by the appropriate eigenvalue.** If a matrix has a linearly independent basis of eigenvectors then we say it is *diagonalizable*. Not all matrices are diagonalizable, but if we choose a matrix at random then it will be (with probability 1), ie. we have to specifically work to construct a matrix that is not diagonalizable. The reason for this is that non-diagonalizable matrices are a low dimensional subset of the space of all matrices. Many arguments in linear algebra are best understood by understanding them for diagonalizable matrices and then generalizing them to the non-diagonalizable case.

The right and left eigenvector equations are given by

$$\lambda v = Av, \qquad \lambda w^T = w^T A \tag{64}$$

respectively. Suppose the columns of $P \in \mathbb{R}^{n \times n}$ are a linearly independent set of right eigenvectors of A and with eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $D \in \mathbb{R}^n$ be a diagonal matrix with the eigenvalues on the diagonal, ie. $D = \mathrm{diag}([\lambda_1, \ldots, \lambda_n])$. The columns of P being right eigenvectors is equivalent to the equation

$$AP = PD \tag{65}$$

$$\Rightarrow \quad A = PDP^{-1} \tag{66}$$

We say that the matrix of eigenvectors P diagonalizes A because it relates A to a diagonal matrix D via a similarity transform. In other words if $x = Pz$, $z' = Px'$ and $x' = Ax$, then $z' = Dz$. Note that in the z-coordinates, D simply scales each coordinate by the appropriate eigenvalue.

Left multiplying (66) by P^{-1} gives $P^{-1}A = DP^{-1}$. Note that this means that the rows of P^{-1} are a set of linearly independent left-eigenvectors of A. Note that this also shows why the left and right eigenvectors come in pairs and share eigenvalues. To summarize, let

$$P = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}, \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix}, \qquad P^{-1} = \begin{bmatrix} - & w_1^* & - \\ & \vdots & \\ - & w_n^* & - \end{bmatrix}, \qquad (67)$$

with v_i and w_j being right and left eigenvectors. A can be decomposed as

$$A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} - & w_1^* & - \\ & \vdots & \\ - & w_n^* & - \end{bmatrix} = \sum_i \lambda_i v_i w_i^* \qquad (68)$$

Note that real eigenvalues denote how much each eigenvectors get stretched when they are multiplied by the matrix.

Computing Eigenvalues and Eigenvectors

As stated above the determinant of a matrix is equal to the product of its eigenvalues. This means that if a matrix has a zero eigenvalue than its determinant is zero. Any vector in the nullspace of a matrix is an eigenvector with an eigenvalue of 0. Note that if $\lambda v = Av$ then $(\lambda I - A)v = 0$. In other words, if v is eigenvector of A with eigenvalue λ, then v is also an eigenvector of $\lambda I - A$ with eigenvalue 0. We can find eigenvalues of A by finding values of λ such that $(\lambda I - A)$ has a 0 eigenvalue. This leads us to characterize eigenvalues as solutions to the equation

$$\chi_A(s) = \det(sI - A) = 0 \qquad (69)$$

$\chi_A(s)$ is called the *characteristic polynomial* of A.

$$\chi_A(s) = \det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$$

Based on properties of determinants, $\chi_A(s)$ will always have order n and the first term will always be s^n.

Once we find roots of $\chi_A(s)$, λ_i, we find the corresponding right and left eigenvectors by finding vectors in the right and left nullspace of $\lambda_i I - A$ respectively.

49

Formulas

2×2 Matrices

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} m+h & p-k \\ p+k & m-h \end{bmatrix}$$

where $m = \frac{1}{2}(a+d)$, $h = \frac{1}{2}(a-d)$, $p = \frac{1}{2}(b+c)$, and $k = \frac{1}{2}(c-b)$

- **Eigenvalues:**

$$\begin{aligned} \lambda_{1,2} &= \frac{\mathrm{Tr}(A)}{2} \pm \sqrt{\left(\frac{\mathrm{Tr}(A)}{2}\right)^2 - \det(A)} \\ &= m \pm \sqrt{h^2 - bc} \\ &= m \pm \sqrt{h^2 + p^2 - k^2} \end{aligned}$$

- **Eigenvectors**

Spectral Mapping Theorem

Polynomial Functions

As stated above computing eigenvectors and eigenvalues simplifies matrix computations. In particular, note that given a diagonalization of $A = PDP^{-1}$, we can compute powers of A as

$$A^k = \underbrace{A \times \cdots \times A}_{\times k} = PD^k \underbrace{P^{-1} \times P}_{I} D^k P^{-1} \times \cdots \times PDP^{-1} = PD^k P^{-1} \tag{70}$$

This implies that if a function $f : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$ is a polynomial (or more generally analytic function) of A, then

$$f(A) = Pf(D)P^{-1} = P \begin{bmatrix} f(\lambda_1) & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & f(\lambda_n) \end{bmatrix} P^{-1} \tag{71}$$

In other words, we can compute polynomial functions of A simply by applying that function to the eigenvalues of A and leaving the eigenvectors unchanged. This is known as the *spectral mapping theorem*. Note that this analysis applies to polynomials with an infinite number of terms such as Taylor expansions of functions such as $e^{(\cdot)}$, $\cos(\cdot)$, and $\sin(\cdot)$ as well.

Matrix Exponential

One important function of A that we want to compute is the *matrix exponential* e^A where which can be defined by its Taylor expansion.

$$e^A := I + A + \tfrac{1}{2!}A^2 + \tfrac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \tfrac{1}{k!}A^k \tag{72}$$

Note that by the spectral mapping theorem we have that

$$e^A = Pe^D P^{-1} = P \begin{bmatrix} e^{\lambda_1} & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & e^{\lambda_n} \end{bmatrix} P^{-1} \tag{73}$$

Exponential functions are interesting because they are functions who are equal to their own derivative (times some scaling), ie. $\frac{d}{dt}e^{\lambda t} = \lambda e^{\lambda t}$. (Note that $e^{\lambda t}$ is actually an *eigenfunction* of the derivative operator $\frac{d}{dt}$ with eigenvalue λ.)

Cayley-Hamilton Theorem

The Cayley-Hamilton theorem says that a matrix satisfies its own characteristic polynomial, ie. $\chi_A(A) = 0$. For diagonalizable matrices, this is a direct application of the spectral mapping theorem.

$$\chi_A(A) = P\chi_A(D)P^{-1} = P \begin{bmatrix} \chi_A(\lambda_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \chi_A(\lambda_n) \end{bmatrix} P^{-1} = 0$$

Consequently,

$$A^n = -\alpha_{n-1}A^{n-1} - \cdots - \alpha_1 A - \alpha_0 I$$

As a result of this, any polynomial function of A could be expressed in terms of powers of A only up through $n-1$. Higher powers of A can be reduced by iteratively plugging in the above equation.

Another application of Cayley-Hamilton gives a polynomial expression for a matrix inverse.

$$0 = \left(A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I \right) A^{-1}$$
$$A^{-1} = -\tfrac{1}{\alpha_0}A^{n-1} - \tfrac{\alpha_{n-1}}{\alpha_0}A^{n-2} - \cdots - \tfrac{\alpha_1}{\alpha_0}I$$

Jordan Form

To motivate a study of Jordan form, we consider the following matrix

$$J_i = \lambda_i I + N_i = \begin{bmatrix} \lambda_i & 1 & \cdots & \cdots & 0 \\ 0 & \lambda_i & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}$$

where N_i is a matrix with 1's on the first super diagonal. This matrix N_i is an example of a *nilpotent matrix* since raising it to some power gives a matrix of 0's, ie. for example

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

Note that any matrix similar to a nilpotent matrix is also nilpotent. If $N_i^k = 0$, then $(PN_iP^{-1})^k = PN_i^kP^{-1} = 0$. If $J_i = \lambda_i I + N_i$, then clearly, $J_i - \lambda_i I$ is nilpotent, ie. $J_i - \lambda_i I = N_i$. Since the eigenvalues of a triangular matrix are just the diagonal values, we have that the only eigenvalue of N_i is simply 0. However, N_i clearly has $n-1$ linearly independent columns, ie. rank $n-1$. Thus it only has a one dimensional nullspace. One can check that the characteristic polynomial of N_i is $\chi_{N_i}(s) = s^n$ and the charactestic polynomial of $J_i = \lambda_i I + N_i$ is $\chi_{J_i}(s) = (s - \lambda_i)^n$.

A matrix is not diagonalizable when a full basis of eigenvectors does not exist. For a matrix $A \in \mathbb{R}^{n \times n}$ with n distinct eigenvalues, there must be a basis of n linearly independent eigenvectors since each eigenvalue λ_i is associated with the nullspace of $\lambda_i I - A$. We know these eigenvectors are linearly independent since if not

$$v_i = \sum_{j \neq i} \alpha_j v_j$$

$$Av_i = A \left(\sum_{j \neq i} \alpha_j v_j \right)$$

$$0 = \sum_{j \neq i} \alpha_j \lambda_j v_j - \lambda_i v_i$$

$$0 = \sum_{j \neq i} \alpha_j (\lambda_j - \lambda_i) v_j$$

An inductive argument shows that $\lambda_i = \lambda_j$ for some i and j which is a contradiction.

In this case, the characteristic polynomial is

$$\chi_A(s) = (s - \lambda_1)(s - \lambda_2) \cdots (s - \lambda_n)$$

In the general case with repeated eigenvaleus, the characteristic polynomial is given by

$$\chi_A(s) = \prod_{i=1}^{k}(s - \lambda_i)^{k_i}$$

where k is the number of distinct eigenvalues and k_i is the number of times each eigenvalue is repeated. If $\dim\big(\mathcal{N}(\lambda_i I - A)\big) = k_i$ for all i, then the matrix is diagonalizable. In this case,

$$\mathcal{N}(\lambda_i I - A) = \mathcal{N}\big((\lambda_i I - A)^2\big) = \mathcal{N}\big((\lambda_i I - A)^3\big) = \dots$$

and

$$\dim\big(\mathcal{N}(\lambda_i I - A)\big) = \dim\big(\mathcal{N}\big((\lambda_i I - A)^2\big)\big) = \dim\big(\mathcal{N}\big((\lambda_i I - A)^3\big)\big) = \dots = k_i$$

This happens when $\mathcal{N}(\lambda_i I - A) \cap \mathcal{R}(\lambda_i I - A) = 0$ for all i.

It is also possible that $\dim\big(\mathcal{N}(\lambda_i I - A)\big) < k_i$ In this case

$$\mathcal{N}(\lambda_i I - A) \subset \mathcal{N}\big((\lambda_i I - A)^2\big) \subset \mathcal{N}\big((\lambda_i I - A)^3\big) \subset \dots$$

and

$$\dim\big(\mathcal{N}(\lambda_i I - A)\big) < \dim\big(\mathcal{N}\big((\lambda_i I - A)^2\big)\big) < \dim\big(\mathcal{N}\big((\lambda_i I - A)^3\big)\big) < \dots < k_i \qquad (74)$$

ie., $\mathcal{N}(\lambda_i I - A) \cap \mathcal{R}(\lambda_i I - A) \neq 0$. A regular eigenvector satisfies

$$(\lambda_i I - A)v_i = 0$$

If $\dim\big(\mathcal{N}(\lambda_i I - A)\big) < \dim\big(\mathcal{N}(\lambda_i I - A)^2\big)$, then we should be able to find generalized eigenvectors that satisfy

$$(\lambda_i I - A)w_i^2 \in \mathcal{N}(\lambda_i I - A), \qquad (\lambda_i I - A)w_i^3 \in \mathcal{N}(\lambda_i I - A)^2, \qquad \text{etc}$$

$w_i^2 \in \mathbb{C}^n$ is a 2nd order eigenvector, $w_i^3 \in \mathbb{C}^n$ is a 3rd order eigenvector, etc.

Note that

$$(\lambda_i I - A)^2 w_i^2 = 0, \qquad (\lambda_i I - A)^3 w_i^3 = 0, \qquad \text{etc}$$

If we are careful in picking, $v_i, w_i^2, w_i^3, \dots$ we can choose them so that

$$0 = (\lambda_i I - A)v_i, \qquad v_i = (\lambda_i I - A)w_i^2, \qquad w_i^2 = (\lambda_i I - A)w_i^3, \qquad \text{etc} \qquad (75)$$

A general organization of these equations is given by

$$AP = PJ = \underbrace{\begin{bmatrix} V_1 & \cdots & V_q \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} J_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q \end{bmatrix}}_{J}$$

53

where

$$V_i = \begin{bmatrix} | & | & | \\ v_1 & w_1^2 & w_1^3 & \cdots \\ | & | & | \end{bmatrix}, \qquad J_i = \lambda_i I + N_i = \begin{bmatrix} \lambda_i & 1 & \cdots & \cdots & 0 \\ 0 & \lambda_i & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}$$

$$N_i = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}$$

J_i is called a Jordan block and q is the number of Jordan blocks. Each Jordan block corresponds to one true eigenvector and a chain of generalized eigenvectors as in (75). Note that if each distinct eigenvalue has only one Jordan block (and only one true eigenvector), then $q = k$, the number of distinct eigenvalues. It is possible that a distinct eigenvalue has more than one Jordan block. In this case, $q > k$. Most matrices are diagonalizable, but every matrix can be put in *Jordan form*. Note that

$$\begin{aligned} A - \lambda_1 I &= PJP^{-1} - \lambda_1 PP^{-1} \\ &= P(J - \lambda_1 I)P^{-1} \\ &= P \begin{bmatrix} N_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q \end{bmatrix} \end{aligned}$$

and that

$$(A - \lambda_1 I)^\ell = P \begin{bmatrix} N_1^\ell & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q^\ell \end{bmatrix}$$

Since N_1 is nilpotent, as ℓ increases the nullspace of $(A - \lambda_1 I)^\ell$ grows as in (74).

We now perform several manipulations with a simple non-diagonalizable matrix to illustrate

some simple properties of Jordan form. Consider

$$
\begin{aligned}
A &= \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix}^{-1} \\
&= \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} - & (q^3)^T & - \\ - & (q^2)^T & - \\ - & p^T & - \end{bmatrix} \\
&= \lambda v (q^3)^T + (v + \lambda w^2)(q^2)^T + (w^2 + \lambda w^3)p^T \\
&= \lambda v (q^3)^T + \lambda w^2 (q^2)^T + \lambda w^3 p^T + v(q^2)^T + w^2 p^T
\end{aligned}
$$

Note that

- The first order right eigenvector v matches up with the third order left generalized eigenvector $(q^3)^T$

- The second order right eigenvector w^2 matches up with the second order left generalized eigenvector $(q^2)^T$

- The third order right eigenvector w^3 matches up with the first order left eigenvector p^T

We note that we could also write A in other ways related to Jordan form (These are just a sample of how the Jordan block and eigenvectors could be shuffled.)

$$
\begin{aligned}
A &= \begin{bmatrix} | & | & | \\ w^2 & v & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 0 & 1 \\ 1 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} - & (q^2)^T & - \\ - & (q^3)^T & - \\ - & p^T & - \end{bmatrix} \\
&= \begin{bmatrix} | & | & | \\ w^3 & w^2 & v \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{bmatrix} \begin{bmatrix} - & p^T & - \\ - & (q^2)^T & - \\ - & (q^3)^T & - \end{bmatrix} \\
&= \text{etc...}
\end{aligned}
$$

6.1 Ordinary Differential Equations (ODEs) and Vector Fields

We model time evolution of a system with *ordinary differential equations* (ODEs). Let $x \in \mathbb{R}^n$ be the vector valued *state* of some system that changes with time. A differential equation is written

$$
\frac{\partial x(t)}{\partial t} = \dot{x}(t) = f(x(t)), \qquad x(0) = x_0, \tag{76}
$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a vector valued function of x and $x_0 \in \mathbb{R}^n$ is some initial state or *initial condition* of the system. If the system has a control input $u(t) \in \mathbb{R}^m$, some signal that we get to choose over time to modify the system dynamics, we can write

$$
\dot{x}(t) = f(x(t), u(t)), \qquad x(0) = x_0, \tag{77}
$$

55

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is now a vector-valued function of x and u. We often refer to differential equations of this form as *vector fields* cause we think of $f(x)$ as defining an evolution direction at each different point x in the state space. We can represent this graphically as a "field" of arrows. $x(t)$ evolves forward in time along these arrows from the initial state x_0 to form a *state trajectory*.

The best understood differential equations are *linear differential equations* that have the form

$$\dot{x} = Ax, \qquad x(0) = x_0 \tag{78}$$

Linear differential equations are nice in that we can compute one *state transition matrix* or state evolution matrix that allows us $x(t)$ for *any initial condition*. In the simple case of (**??**) this transition matrix is given by e^{At} and we have that

$$x(t) = e^{At}x(0) \tag{79}$$

If A is diagonalizable, ie. $A = PDP^{-1}$ for eigenbasis P then by the spectral mapping theorem, we have that $e^{At} = Pe^{Dt}P^{-1}$. If we represent x in the eigenvector coordinates, ie. $x(t) = Pz(t)$ or $z(t) = P^{-1}x(t)$ this equation becomes

$$z(t) = e^{Dt}z(0) \qquad \rightarrow \qquad \begin{bmatrix} z_1(t) \\ \vdots \\ z_n(t) \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 t}z_1(t) \\ \vdots \\ e^{\lambda_n t}z_n(t) \end{bmatrix} \tag{80}$$

ie. we can examine the evolution of each eigenvector separately. Suppose $\lambda = a + bi$. The exponential $e^{\lambda t} = e^{(a+bi)t} = e^{at}e^{bti}$. Thus a controls the decay (or explosion) rate of the signal and b (referred to as the *frequency*) determines the oscillation rate of the signal.

First, suppose all the eigenvalues of a real matrix A are real. The possible vector fields and state trajectories (for $A \in \mathbb{R}^{2 \times 2}$ are illustrated below.

7 Rotational Motion

7.1 Rotation Matrices

A *rotation* matrix is a real valued matrix whose determinant is 1, and whose columns are *orthonormal*, ie. whose columns are orthogonal and have length 1. Note that these conditions can be succinctly written as

$$\det(R) = 1, \qquad R^T R = I \tag{81}$$

This second condition is that the left-inverse of R is its transpose, ie. $R^T = R^{-1}$. Note that this also means that $RR^{-1} = RR^T = I$, ie. that the rows of R are orthonormal as well. The generalization of rotation matrices to complex matrices are called *unitary* matrices, ie. $U \in \mathbb{C}^{n \times n}$

that satisfy $\det(U) = 1$ and $U^*U = I$. The columns of a rotation matrix often used to represent an orthonormal coordinate system for $\mathbb{R}^n$. An orthonormal coordinate transformation is a type of *isometry*, a coordinate transformation that does not change metric of the space. One can see this by considering the coordinate transform $x = Rz$. All metric properties (distances and angles) are computed using inner products. Note that $z^Tz = x^TR^TRx = x^Tx$. As a result inner products in the x or z coordinates are the same so no metric properties change by transforming between these coordinate systems. This is consistent with our intuition about rotation matrices.

Relaxing the determinant restriction to $\det(R) = \pm1$ allows the set of matrices to include rotations and reflections. This allows for left-handed as well as right-handed coordinate systems. A matrix R such that $R^TR = I$, $\det R = -1$ represents a left-handed coordinate system instead of a right-handed one. In order to transform from the standard basis coordinates to a left-handed coordinate system at least one reflection is required. Volumes transformed by orthogonal matrices such that $\det R = -1$ are flipped inside out.

7.2   Skew-Symmetric Matrices

A matrix $K \in \mathbb{R}^{n \times n}$ is *skew-symmetric* if $K = -K^T$. Skew-symmetric matrices have purely imaginary eigenvalues. Best thought of in the context of the equation $\dot{x} = Kx$. Note that you can prove explicitly that if $K = -K^T$ then $x^TKx = 0$, ie. Kx is perpendicular to x. Vector field rotational, ie. e^{Kt} is a rotation matrix. Eigenvalues come in complex conjugate pairs, so if you have an odd dimensional matrix then there is always one such that $\lambda = 0$. It follows that $e^{\lambda t} = e^{0t} = 1$ and the eigenvector associated with the 0 eigenvalue is the axis of rotation.

Skew-Symmetric Matrices / Rotational Vector Fields

2D

3D

$$\dot{x} = Kx$$

2D:
$$K = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

$$e^{Kt} = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

3D:
$$K = \begin{bmatrix} v_1 & v_2 & \omega \end{bmatrix} \begin{bmatrix} 0 & -b & 0 \\ b & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \omega^T \end{bmatrix}$$

$$e^{Kt} = \begin{bmatrix} v_1 & v_2 & \omega \end{bmatrix} \begin{bmatrix} \cos(bt) & -\sin(bt) & 0 \\ \sin(bt) & \cos(bt) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \omega^T \end{bmatrix}$$

7.3 Matrix Lie Groups

Matrices can be thought of both as *vectors* or as *operators*. Sets of matrices are sometimes endowed with a local coordinate structure. These sets of matrices can be thought of as a *manifold* allowing one to talk about surface or space like properties of this set of matrices along with curvature and other geometric notions. Matrices as operators often represent transformations that encode symmetries, such as permutations, reflections, rotations, etc. In this context we can often talk about sets of matrices as mathematical *groups*. Group theory is interested in composing operations or transformations and seeing whether or not the composite operation has the same properties as the original two. In some case in the contexts of matrices, a set of matrices is both a manifold and a group. Perhaps the best example of these sets of matrices is the set of rotation matrices. Rotation matrices have a continuous manifold structure but also a clear set of symmetries that are preserved when two rotations are multiplied together. These matrix groups with a manifold structure are called *Lie groups*. Some well known examples are

- $GL(n)$: General linear group of dimension n, $n \times n$ matrices with $\det(A) = 1$ for $A \in GL(n)$.

  $\Rightarrow$) Lie algebra: $GL(n)$.

- $SO(n)$: Special orthogonal group of matrices $R \in \mathbb{R}^{n \times n}$ such that $R^T R = I$ and $\det R = 1$. This is the group of rotation matrices.

  $\Rightarrow$) Lie algebra: skew symmetric matrices.

- $U(n)$: Unitary group of matrices $U \in \mathbb{C}^{n \times n}$ such that $U^T U = I$.

  $\Rightarrow$) Lie algebra: skew Hermitian matrices.

Rotation matrices are an excellent example of a Lie group. If $R_1, R_2 \in SO(n)$ then clearly $R_1 R_2 \in SO(n)$. In addition there exists a way to perturb a rotation matrix so that the new matrix

is still a rotation, ie. there is some locally "flat" (Euclidean) structure to the space of matrices. This allows us to think of objects that lie in the tangent space to this manifold as infinitesimal rotations. The tangent space to the Lie group at the identity defines the *Lie algebra* of the Lie group. Integrating over an element in the Lie algebra builds up a particular element in the group. The Lie algebra for $SO(n)$ is the space of skew-symmetric matrices and the equation

$$R = e^{Kt}, \qquad \text{for} \quad K = -K^T$$

is an example of this integration. Geometrically, this concept can be visualizad as follows.

7.4 Complex Eigenvalues and Eigenvectors

Along with representing stretching, complex eigenvalues can represent rotation of vectors as well. If a real matrix has complex eigenvalues then they come in complex conjugate pairs. The eigenvectors come in conjugate pairs as well. We detail the exact mechanics of this below. Consider a matrix A with the first two eigenvalues and left and right eigenvectors given as

$$\lambda_1 = a + bi, \qquad r_1 = \tfrac{1}{\sqrt{2}}\big(u + vi\big), \qquad \ell_1 = \tfrac{1}{\sqrt{2}}\big(w + yi\big) \tag{82}$$

$$\lambda_2 = a - bi, \qquad r_2 = \tfrac{1}{\sqrt{2}}\big(u - vi\big), \qquad \ell_2 = \tfrac{1}{\sqrt{2}}\big(w - yi\big) \tag{83}$$

with real vectors u, v, w, y. The diagonal form of the matrix is given by

$$
A = \begin{bmatrix} | & | & \\ r_1 & r_2 & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} - & \ell_1^* & - \\ - & \ell_2^* & - \\ & \vdots & \end{bmatrix}
\tag{84}
$$

$$
= \begin{bmatrix} | & | & \\ \frac{1}{\sqrt{2}}(u+vi) & \frac{1}{\sqrt{2}}(u-vi) & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} a+bi & 0 & \cdots & 0 \\ 0 & a-bi & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} - & \frac{1}{\sqrt{2}}(w-yi)^T & - \\ - & \frac{1}{\sqrt{2}}(w+yi)^T & - \\ & \vdots & \end{bmatrix}
\tag{85}
$$

$$
= \begin{bmatrix} | & | & \\ u & v & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} a & b & \cdots & 0 \\ -b & a & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} - & w^T & - \\ - & y^T & - \\ & \vdots & \end{bmatrix}
\tag{86}
$$

This last equality is not obvious so we detail it below, but first note the form of the diagonal block, with the real parts of the eigenvalues on the diagonal and the imaginary parts on the off diagonal. Define the 2×2 complex matrix

$$
U = \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}.
\tag{87}
$$

Note that U is a unitary matrix, ie. $UU^* = I$. The first two dyads of the diagonalization are given by

$$
\sum_{i=1,2} \lambda_i r_i \ell_i^* = \begin{bmatrix} | & | \\ \frac{1}{\sqrt{2}}(u-vi) & \frac{1}{\sqrt{2}}(u+vi) \\ | & | \end{bmatrix} \begin{bmatrix} a+bi & 0 \\ 0 & a-bi \end{bmatrix} \begin{bmatrix} - & \frac{1}{\sqrt{2}}(w+yi)^T & - \\ - & \frac{1}{\sqrt{2}}(w-yi)^T & - \end{bmatrix}
\tag{88}
$$

$$
= \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \tfrac{1}{\sqrt{2}}}_{U} \underbrace{\tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}}_{U^*} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \tfrac{1}{\sqrt{2}}}_{U} \underbrace{\tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}}_{U^*} \begin{bmatrix} -w^T- \\ -y^T- \end{bmatrix}
\tag{89}
$$

$$
= \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} -w^T- \\ -y^T- \end{bmatrix}
\tag{90}
$$

Note that both

$$\begin{bmatrix} - & w^T & - \\ - & y^T & - \end{bmatrix} \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad \frac{1}{2} \begin{bmatrix} - & (w+yi)^T & - \\ - & (w-yi)^T & - \end{bmatrix} \begin{bmatrix} | & | \\ u-vi & u+vi \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$(91)$$

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} - & w^T & - \\ - & y^T & - \end{bmatrix} \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \frac{1}{\sqrt{2}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$(92)$$

We could write the eigenvalues as well in polar form $\lambda_1 = a + bi = \gamma e^{i\theta}$ and $\lambda_2 = a - bi = \gamma e^{-i\theta}$ where $\gamma = |\lambda_1| = |\lambda_2| = \sqrt{a^2 + b^2} \geq 0$. If we consider the diagonal block associated with the complex eigenvectors we have that

$$\begin{bmatrix} a+bi & 0 \\ 0 & a-bi \end{bmatrix} = \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \tag{93}$$

$$\gamma \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} = \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \gamma \underbrace{\begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix}}_{R} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \tag{94}$$

remembering that $e^{i\theta} = c\theta + s\theta i$. Note that R is rotation matrix. From this form, we can see that complex eigenvalues of a real matrix correspond to rotations and stretching. The stretching parameter is given by γ, the rotation is given by the matrix R and the corresponding angle θ, and the plane of rotation is the subspace spanned by the vectors u and v. The relationship of the vectors u and v (particularly the value of the inner product $u^T v$) determines the shape of the rotation. If u and v are orthogonal ($u^T v = 0$), then the rotation is circular. If not, then the rotation has an ellipsoidal shape.

Note that right and left eigenvector pairs are not unique but can be scaled by either a real or complex value. This is because the equation $\lambda v = Av$ does not specify the length of v. In diagonal form, since diagonal matrices commute, we can write (shown in here in the 2×2 case, but easily extended)

$$\begin{bmatrix} | & | \\ r_1 & r_2 \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} - & \ell_1^* & - \\ - & \ell_n^* & - \end{bmatrix} = \begin{bmatrix} | & | \\ r_1 & r_2 \\ | & | \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \alpha_1^{-1} & 0 \\ 0 & \alpha_2^{-1} \end{bmatrix} \begin{bmatrix} - & \ell_1^* & - \\ - & \ell_2^* & - \end{bmatrix}$$

$$(95)$$

$$= \begin{bmatrix} | & | \\ \alpha_1 r_1 & \alpha_2 r_2 \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} - & \alpha_1^{-1}\ell_1^* & - \\ - & \alpha_2^{-1}\ell_2^* & - \end{bmatrix} \tag{96}$$

Thus we can scale the right eigenvectors by α_i and the left eigenvectors by α_i^{-1} and the diagonal form of the matrix stays the same. Note that in general α_i can be either real or complex. One

interesting special case is the case of a real matrix with complex eigenvalues where $\alpha = e^{i\phi}$

$$\begin{bmatrix} | & | \\ \frac{1}{\sqrt{2}}(u - vi) & \frac{1}{\sqrt{2}}(u + vi) \\ | & | \end{bmatrix} \begin{bmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{bmatrix} \begin{bmatrix} \gamma e^{i\theta} & 0 \\ 0 & \gamma e^{-\theta i} \end{bmatrix} \begin{bmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{bmatrix} \begin{bmatrix} - & \frac{1}{\sqrt{2}}(w + yi)^T & - \\ - & \frac{1}{\sqrt{2}}(w - yi)^T & - \end{bmatrix} =$$

$$= \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} c\phi & -s\phi \\ s\phi & c\phi \end{bmatrix} \gamma \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix} \begin{bmatrix} c\phi & s\phi \\ -s\phi & c\phi \end{bmatrix} \begin{bmatrix} - & w^T & - \\ - & y^T & - \end{bmatrix} \tag{97}$$

$$= \begin{bmatrix} | & | \\ c\phi u + s\phi v & -s\phi u + c\phi v \\ | & | \end{bmatrix} \gamma \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix} \begin{bmatrix} - & c\phi w^T + s\phi y^T & - \\ - & -s\phi w^T + c\phi y^T & - \end{bmatrix} \tag{98}$$

$$= \begin{bmatrix} | & | \\ u' & v' \\ | & | \end{bmatrix} \gamma \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix} \begin{bmatrix} - & w'^T & - \\ - & y'^T & - \end{bmatrix} \tag{99}$$

with $u' = c\phi u + s\phi v$, $v' = -s\phi u + c\phi v$, $w' = c\phi w + s\phi y$, $y' = -s\phi w + c\phi y$.

Note that from this analysis, we can see that the vectors u and v (and also w and y) are not unique but can be rotated by some phase ϕ. This can be derived either from the fact that diagonal matrices commute or from the fact that 2-D rotation matrices commute. Indeed, computing the quantity

$$\begin{bmatrix} | & | \\ u' & v' \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ c\phi u + s\phi v & -s\phi u + c\phi v \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} c\phi & s\phi \\ -s\phi & c\phi \end{bmatrix} \tag{100}$$

is equivalent to changing the basis vectors that span the subspace containing u and v. It is not obvious but a more complicated analysis (not shown) gives that u' and v' give the same ellipsoidal shape of rotation as u and v.

8 Symmetric Matrices

8.1 Symmetric and Hermitian Matrices

A *symmetric* matrix is a real matrix $S \in \mathbb{R}^{n \times n}$ such that $S = S^T$. A Hermitian matrix $H \in \mathbb{C}^{n \times n}$ is a matrix such that $H = H^*$. Symmetric matrices have all real eigenvalues and can be diagonalized by rotation matrices, ie. for every symmetric matrix S, there exists a rotation matrix R and a diagonal matrix of real eigenvalues D such that

$$S = RDR^T = \begin{bmatrix} | & & | \\ r_1 & \cdots & r_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} - & r_1^T & - \\ & \vdots & \\ - & r_n^T & - \end{bmatrix} \tag{101}$$

This means that multiplying by an $n \times n$ symmetric matrix corresponds to stretching in n orthogonal coordinate directions. Symmetric matrices can be thought of as defining level sets of the quadratic form $h(x) = \frac{1}{2}x^T S x$. The vector field $\dot{x} = Sx$ is then a *gradient field*, $\dot{x} = \frac{\partial h}{\partial x}^T$. Intuitively, $\dot{x} = Sx$ can be thought of as flowing up a surface defined by $h(x)$. This is also called a *conservative* vector field in physics and the function $h(x)$ is typically related to the energy of a system (or some analog). The general condition for a nonlinear vector field $\dot{x} = f(x)$ to be conservative, ie. that $f(x) = \frac{\partial h}{\partial x}^T$ for some $h(x)$ is given by $\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}$ for all i, j. This condition is derived from the fact that if such and $h(x)$ exists, then $\frac{\partial^2 h}{\partial x_i \partial x_j} = \frac{\partial^2 h}{\partial x_j \partial x_i}$.

8.2 Positive Definiteness

We say a symmetric matrix S is *positive definite* if

$$x^T S x > 0, \qquad \text{for all } x \in \mathbb{R}^n \tag{102}$$

Since we could pick x to be any r_i, it follows that for a positive definite matrix, $\lambda_i > 0$ for all i. If the "greater than" signs above are replaced with "greater than or equal" signs, ie. "$>$" is replaced with "$\geq$" then we say the matrix is *positive semi-definite*. Similarly, if we replace the "greater than" signs with "less than" signs we say that the matrix is *negative definite* or *negative semi-definite*.

8.3 Polar Decomposition

For any matrix $A \in \mathbb{R}^{m \times n}$, there are two positive semi-definite matrices $P = (A^T A)^{\frac{1}{2}}$ and $P' = (AA^T)^{\frac{1}{2}}$ (where the positive square root of each eigenvalue is taken) that are closely related to the "shape" of the matrix A. Similarly to how the magnitude of a complex number is defined by $|z| = \sqrt{z^* z}$, we can say that the "magnitude and shape" of A is defined either by $(A^T A)^{\frac{1}{2}}$ or

$(AA^T)^{\frac{1}{2}}$. Expanding out A^TA, we get

$$A^TA = \begin{bmatrix} - & A_1^T & - \\ & \vdots & \\ - & A_n^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ A_1 & \cdots & A_n \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} A_1^TA_1 & \cdots & A_1^TA_n \\ \vdots & & \vdots \\ A_n^TA_1 & \cdots & A_n^TA_n \end{bmatrix} = \begin{bmatrix} |A_1||A_1|\cos(\theta_{11}) & \cdots & |A_1||A_n|\cos(\theta_{1n}) \\ \vdots & & \vdots \\ |A_n||A_1|\cos(\theta_{n1}) & \cdots & |A_n||A_n|\cos(\theta_{nn}) \end{bmatrix}$$

We note that this matrix is fully determined by the size and relative orientation of the columns of A. Another way to say this is that applying an orthonormal transformation to all the columns of A does not change A^TA. Indeed $(RA)^T(RA) = A^TR^TRA = A^TA$. Similarly the size and relative orientation of the rows of A full determines AA^T. We can make precise the sense in which A has the same shape as $P = (A^TA)^{\frac{1}{2}}$ by noting that P and A differ by a orthonormal transformation. Indeed,

$$A = \underbrace{A(A^TA)^{-\frac{1}{2}}}_{R} \underbrace{(A^TA)^{\frac{1}{2}}}_{P}$$

We note that we can check that $R^TR = I$

$$R^TR = (A^TA)^{-\frac{1}{2}}A^TA(A^TA)^{-\frac{1}{2}}$$

$$= \underbrace{(A^TA)^{-\frac{1}{2}}(A^TA)^{\frac{1}{2}}}_{I}\underbrace{(A^TA)^{\frac{1}{2}}(A^TA)^{-\frac{1}{2}}}_{I} = I$$

and thus we have that rotating (and possibly reflecting) all the columns of the positive semidefinite matrix $P \succeq 0$ by one rotation gives A. Similarly a complex number $z = |z|e^{i\theta}$ can be created by starting with it's norm $|z| \geq 0$ and rotating it in the complex plane by $e^{i\theta}$. Similarly A can be created from $(AA^T)^{\frac{1}{2}}$ by applying the orthonormal transformation $R' = (AA^T)^{-\frac{1}{2}}A$.

$$A = \underbrace{(AA^T)^{\frac{1}{2}}}_{P'} \underbrace{(AA^T)^{-\frac{1}{2}}A}_{R'}$$

One can check that in fact for a square matrix, $R = R'$.

This leads us to the polar decomposition. A square, invertible $A \in \mathbb{R}^{n \times n}$ can be written in a polar form similar to the polar decomposition of a complex number $z = \sqrt{z^*z}e^{i\theta}$. The

$$A = RP = P'R$$

where

$$P = (A^*A)^{\frac{1}{2}} = V\Sigma V^*$$

$$P = (AA^*)^{\frac{1}{2}} = U\Sigma U^*$$

$$R = (AA^*)^{\frac{-1}{2}}A = A(A^*A)^{\frac{-1}{2}} = UV^*$$

Note that $P \succeq 0$ and $P' \succeq 0$. Note also the There are two separate versions of the polar decomposition one with P and one with P'. Note also the connections between the polar decomposition and the singular value decomposition (see below). These relationships can be checked directly.

9 Singular Value Decomposition

The *singular value decomposition* (SVD) provides even more insight beyond the polar decomposition. The SVD is very general and can apply to any matrix $A \in \mathbb{C}^{m \times n}$ even if the matrix is not-invertible, not diagonalizable, or even not square or full-rank. As a result, it is an often used, powerful analysis tool. We will perform the derivation below for $A \in \mathbb{R}^{m \times n}$ but we note that the same derivation works for complex matrices using conjugate transposes.

To construct the SVD, we will assume that A is fat or square and analyze $A^T A$. (A similar construction can be done if A is tall using AA^T.). We first diagonalize $A^T A$ as

$$A^T A = VDV^T$$

where $V \in \mathbb{R}^{n \times n}$ is orthonormal, ie. $V^T V = I$, and $D \succeq 0$ is diagonal, real and positive semi-definite. We know this is possible since $A^T A$ is symmetric and thus has all real eigenvalues and orthogonal eigenvectors. We will also assume that the diagonal of D is arranged in descending order from the largest eigenvalue to the smallest. Since A is fat or square, some portion of the diagonal of D will be 0's. We can take the positive square root of D to get

$$D^{\frac{1}{2}} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix}$$

and the block zeros have the appropriate sizes. $\{\sigma_i\}_{i=1}^{k}$ are called *singular values* and are the positive square roots of the nonzero eigenvalues of $A^T A$. We can enumerate V as

$$V = \begin{bmatrix} V^1 & V^2 \end{bmatrix}$$

where

$$V^1 = \begin{bmatrix} | & & | \\ V_1 & \cdots & V_k \\ | & & | \end{bmatrix}, \qquad V^2 = \begin{bmatrix} | & & | \\ V_{k+1} & \cdots & V_n \\ | & & | \end{bmatrix}$$

where the columns of V^1 correspond to positive, non-zero singular values and the columns of V^2 are the eigenvectors for the zero eigenvalues of $A^T A$. We note that the columns of V^2 can be chosen somewhat arbitrarily as long as they are orthonormal and span $\mathcal{N}(A^T A)$. Since $\mathcal{N}(A^T A) = \mathcal{N}(A)$, $\mathcal{R}(V^2) \in \mathcal{N}(A)$ as well.

We note that any columns corresponding to repeated eigenvalues of $A^T A$ (including repeated zeros), may be arbitrarily chosen (as long as they are orthonormal). We can now define a matrix $U \in \mathbb{R}^{m \times m}$ as

$$U = \begin{bmatrix} U^1 & U^2 \end{bmatrix}$$

where

$$U^1 = \begin{bmatrix} | & & | \\ U_1 & \cdots & U_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \frac{AV_1}{\sigma_1} & \cdots & \frac{AV_k}{\sigma_k} \\ | & & | \end{bmatrix} \qquad U^2 = \begin{bmatrix} | & & | \\ U_{k+1} & \cdots & U_m \\ | & & | \end{bmatrix}$$

Note that the columns of U^1 are orthonormal.

$$\frac{V_i^T A^T}{\sigma_i} \frac{AV_i}{\sigma_i} = \frac{\sigma_i^2}{\sigma_i^2} V_i^T V_i = 1, \qquad \frac{V_i^T A^T}{\sigma_i} \frac{AV_j}{\sigma_j} = \frac{\sigma_i^2}{\sigma_j^2} V_i^T V_j = 0$$

The columns of U_2 can be chosen to complete an orthonormal basis for $\mathbb{R}^m$ and thus $U^T U = I$. We can then write

$$U^1 \Sigma = AV^1$$

By adding $(0)U^2$ to the left-hand side, and since $V^2 \in \mathcal{N}(A)$, we can write

$$\begin{bmatrix} U^1 \Sigma + (0)U^2 & 0 \end{bmatrix} = A \begin{bmatrix} V^1 V^2 \end{bmatrix}$$
$$\begin{bmatrix} U^1 & U^2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} = A \begin{bmatrix} V^1 V^2 \end{bmatrix}$$

right multiplying by V^T gives

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T$$
$$= \begin{bmatrix} U^1 & U^2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^{1T} \\ V^{2T} \end{bmatrix}$$

This is the singular value decomposition. Note the following relationships

$$\mathcal{R}(U^1) = \mathcal{R}(A), \qquad \mathcal{R}(V^1) = \mathcal{R}(A^T)$$
$$\mathcal{R}(U^2) = \mathcal{N}(A^T), \qquad \mathcal{R}(V^2) = \mathcal{N}(A)$$

Note also that the columns of U are orthonormal eigenvectors of AA^T with the singular values squared as eigenvalues. Indeed,

$$AA^T U_i = \frac{AA^T AV_i}{\sigma_i} = \frac{\sigma_i^2}{\sigma_i} AV_i = (\sigma_i)^2 U_i$$

Thus this construction would have worked using AA^T instead of $A^T A$ if desired.

9.1 Symmetric-Skew Symmetric/Helmholtz Decomposition

A square matrix $A \in \mathbb{R}^{n \times n}$ can be decomposed as follows

$$A = \underbrace{\tfrac{1}{2}\left(A + A^T\right)}_{S} + \underbrace{\tfrac{1}{2}\left(A - A^T\right)}_{K}$$

Note that $S = S^T$ is symmetric and $K = -K^T$ is skew-symmetric. This decomposition says that the space of real matrices is actually the direct sum of the space of symmetric matrices and the space of skew-symmetric matrices. Under the vectorized matrix inner product $\langle \cdot, \cdot \rangle = \text{Tr}\left((\cdot)^T(\cdot)\right)$, we have that

$$\langle S, K \rangle = \text{Tr}(S^T K) = \sum_{i,j} S_{ij} K_{ij}$$
$$= \sum_{i<j} S_{ij} K_{ij} + \sum_{i>j} S_{ij} K_{ij} + \sum_{i} S_{ii} K_{ii}$$
$$= \sum_{i<j} S_{ij} K_{ij} + \sum_{j>i} S_{ji} K_{ji} + \sum_{i} S_{ii} K_{ii}$$
$$= \sum_{i<j} S_{ij} \left(K_{ij} + K_{ji}\right) + \sum_{i} S_{ii} K_{ii} = 0$$

If the definition of positive definite is extended to non-symmetric matrices we have that

$$x^T A x = x^T S x + x^T K x = x^T S x$$

and thus A is positive definite if and only if $S = \tfrac{1}{2}(A + A^T)$ is positive definite. In the context of vector fields, this means that any linear vector field $\dot{x} = Ax$ can be decomposed into a conservative piece and a rotational piece.

$$\dot{x} = Ax = \underbrace{Sx}_{\text{conservative}} + \underbrace{Kx}_{\text{rotational}}$$

which is a special application of the Helmholtz decomposition to linear vector fields.

10 Matrix vs. Complex Number Analogies

The polar decomposition and the symmetric/skew-symmetric decomposition provide several deep analogies between matrices and complex numbers. Symmetric matrices act a lot like real numbers and positive definite symmetric matrices act like positive real numbers. Among other things, they have real and positive real eigenvalues respectively. Skew-symmetric matrices act a lot like purely imaginary numbers, encoding rotational flow and having purely imaginary eigenvalues. We can even draw a "complex plane" of sorts for matrices with symmetric matrices as the real axis and skew-symmetric matrices as the vertical axis. As shown above, the fact that these two subspaces are orthogonal is accurate. The Cartesian description of complex numbers is analogous to the symmetric-skew symmetric decomposition and the polar description of complex numbers is analogous to the polar decomposition.

$$z = a + bi, \qquad \Rightarrow \qquad A = S + K$$
$$z = |z|e^{i\theta} \qquad \Rightarrow \qquad A = RP = P'R$$

This analogy extends in the following ways detailed in the diagram. For complex numbers and square invertible matrices and

$$z = a + bi = |z|e^{i\theta}, \qquad A = U\Sigma V^T$$

- $z \Rightarrow A$, $z^* \Rightarrow A^T$, $z^{-1} \Rightarrow A^{-1}$, $z^{-*} \Rightarrow A^{-T}$

- $a \Rightarrow S = \frac{1}{2}(A + A^T)$, $bi \Rightarrow K = \frac{1}{2}(A - A^T)$.

- $|z| = (z^*z)^{\frac{1}{2}} \Rightarrow P = (A^T A)^{\frac{1}{2}} = V\Sigma V^T$, $P' = (AA^T)^{\frac{1}{2}} = U\Sigma U^T$,

- $e^{i\theta} \Rightarrow R = A(A^T A)^{-\frac{1}{2}} = (AA^T)^{-\frac{1}{2}}A = UV^T$, $e^{-i\theta} \Rightarrow R^T$,

skew

$$A = RP = P'R = U \Sigma V^T$$
$$P = (A^T A)^{\frac{1}{2}} = V \Sigma V^T$$
$$P' = (A A^T)^{\frac{1}{2}} = U \Sigma U^T$$
$$R = A(A^T A)^{-\frac{1}{2}} = (A A^T)^{-\frac{1}{2}} A = U V^T$$
$$S = \tfrac{1}{2}(A + A^T)$$
$$K = \tfrac{1}{2}(A - A^T)$$

A

R

K

A^{-T}

S

P or P'

Sym

P^{-1} or $(P')^{-1}$ I

A^{-1}

-k

R^T

A^T

MATRIX DECOMPOSITION
CARTOON

- Sym/skew-sym } ADD.
- SVD
- POLAR DECOMP } MULTI.

11 Homogeneous Tranformations

The group of *homogeneous transformations* - the special Euclidean group $SE(3)$, $SE(3) = SO(3) \times \mathbb{R}^3$ - is the set of rotations and translations in $\mathbb{R}^3$. Just as rotations $SO(3)$ have a matrix representation in $\mathbb{R}^{3\times3}$, so homogeneous transformations have a matrix representation in $\mathbb{R}^{4\times4}$. For $g \in SE(3)$, we can write

$$g = \begin{bmatrix} R & p \\ \mathbf{0} & 1 \end{bmatrix}$$

where R is a rotation, p is a translation vector and $\mathbf{0} = [0, 0, 0]$. Points in space in this representation are represented by a vector p with a 1 concatenated at the bottom. Relative distances or velocities are have a 0 concatenated.

Positions: $\begin{bmatrix} p \\ 1 \end{bmatrix}$, Rel. positions Velocities $\begin{bmatrix} p \\ 1 \end{bmatrix} - \begin{bmatrix} q \\ 1 \end{bmatrix} = \begin{bmatrix} p - q \\ 0 \end{bmatrix}$, $\begin{bmatrix} v \\ 0 \end{bmatrix}$

11.1 Homogeneous Coordinate Transformations

We give the following example to expound homogeneous coordinate transformations. g_{AB} represents the coordinate transformation that takes a point in frame B and transforms it into frame A.

We consider the following coordinate frames.

The transformation

$$g_{AB} = \begin{bmatrix} R_{AB} & p_{AB} \\ \mathbf{0} & 1 \end{bmatrix}$$

is the homogeneous transformation that takes vectors written in the B-frame and transforms them to the A-frame, ie.

$$\begin{bmatrix} q_A \\ 1 \end{bmatrix} = \begin{bmatrix} R_{AB} & p_{AB} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} q_B \\ 1 \end{bmatrix} = \begin{bmatrix} R_{AB}q_B + p_{AB} \\ 1 \end{bmatrix}$$

The columns of the rotation matrix R_{AB} should form an orthonormal (right handed) coordinate system. Since

$$q_A = R_{AB}q_B + p_{AB}.$$

if q_B is the coordinates of the point with respect to the B-frame, then the columns of R_{AB} should be the coordinates of the axes of the B-frame with respect to the A frame. q_A is then the sum of $R_{AB}q_B$ and the translation vector from A to B, p_{AB}.

Using this intuition and the noting the relationships between different axes we can get

$$g_{AB} = \begin{bmatrix} R_{AB} & p_{AB} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad g_{BC} = \begin{bmatrix} R_{BC} & p_{BC} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$g_{CD} = \begin{bmatrix} R_{CD} & p_{CD} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 3 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We can compose these three transformations to get g_{AD}.

$$g_{AD} = g_{AB}g_{BC}g_{CD}$$
$$= \begin{bmatrix} R_{AB} & p_{AB} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} R_{BC} & p_{BC} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} R_{CD} & p_{CD} \\ \mathbf{0} & 1 \end{bmatrix}$$
$$\begin{bmatrix} R_{AD} & p_{AD} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} R_{AB}R_{BC}R_{CD} & p_{AB} + R_{AB}p_{BC} + R_{AB}R_{BC}p_{CD} \\ \mathbf{0} & 1 \end{bmatrix}$$

$$R_{AD} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Notice that each column of R_{AD} is the coordinates of an axis of the D-frame in the A-frame.

$$p_{AB} + R_{AB}p_{BC} + R_{AB}R_{BC}p_{CD} = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{2}} \\ 6 \\ -\frac{3}{\sqrt{2}} \end{bmatrix}$$

$$g_{AD} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} & \frac{3}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 6 \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{1}{2} & -\frac{3}{\sqrt{2}} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The transformation g_{DA} is then given by $g_{DA} = g_{AD}^{-1}$.

$$g_{DA} = g_{AD}^{-1} = \begin{bmatrix} R_{AD}^T & -R_{AD}^T p_{AD} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} R_{DA} & -R_{DA}p_{AD} \\ \mathbf{0} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{3}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{9}{\sqrt{2}} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

11.2   Twist Motions

Homogeneous transformations are generated by twist motions represented by $\xi = [v^T, \omega^T]$ where v is a velocity and ω is an axis and magnitude of rotation. An element in the Lie algebra of $SE(3)$, denoted $se(3)$, has a matrix representation given by $\hat{\xi} \in se(3)$ as

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ \mathbf{0} & 0 \end{bmatrix}$$

When we define $\hat{\xi}$, we are thinking of a homogeneous transformation as the state transition matrix for a differential equation

$$\dot{x} = \hat{\omega}x + v$$

$$\begin{bmatrix} \dot{x} \\ 1 \end{bmatrix} = \begin{bmatrix} \hat{\omega} & v \\ \mathbf{0} & 0 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

Solution: $\Rightarrow$ $\begin{bmatrix} x(t) \\ 1 \end{bmatrix} = g(t) \begin{bmatrix} x(0) \\ 1 \end{bmatrix} = e^{\hat{\xi}t} \begin{bmatrix} x(0) \\ 1 \end{bmatrix}$

Just as $R(t) = e^{\hat{\omega}t} \in SO(3)$, we have that $g(t) = e^{\hat{\xi}t} \in SE(3)$.

General formulas for integrating $\hat{\xi}$ to get g are given by

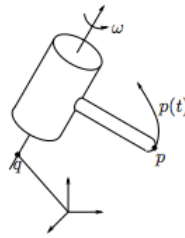$$\text{Case 1: } \omega = 0 \qquad g(t) = \begin{bmatrix} I & vt \\ \mathbf{0} & 1 \end{bmatrix}$$

$$\text{Case 2: } \omega \neq 0 \qquad g(t) = \begin{bmatrix} e^{\hat{\omega}t} & (I - e^{\hat{\omega}t})\hat{\omega}v + \omega\omega^T vt \\ \mathbf{0} & 1 \end{bmatrix}$$

11.3 Manipulator Transformations

11.3.1 Joint Motions

Joint motions of a robotic manipulator can be represented using twists

- **Revolute Joint:** A revolute joint rotates a point p around an axis ω through a point q.

$$\dot{p} = \hat{\omega}(p - q)$$
$$= \hat{\omega}p - \hat{\omega}q$$

Thus the twist for a revolute joint is given by $\xi = [(-\hat{\omega}q)^T \ \omega^T]^T$. For a rotation by an angle θ about the axis we get a transformation of

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & -\hat{\omega}q \\ \mathbf{0} & 0 \end{bmatrix}, \qquad e^{\hat{\xi}\theta} = \begin{bmatrix} e^{\hat{\omega}\theta} & (I - e^{\hat{\omega}\theta})\,\hat{\omega}\big(-\hat{\omega}q\big) - \omega\omega^T\hat{\omega}q\theta \\ \mathbf{0} & 1 \end{bmatrix}$$
$$= \begin{bmatrix} e^{\hat{\omega}\theta} & -(I - e^{\hat{\omega}\theta})\,\hat{\omega}^2 q \\ \mathbf{0} & 1 \end{bmatrix}$$

- **Prismatic Joint:**

  A prismatic joint slides points linearly.

$$\dot{p} = v$$

Thus the twist for a prismatic joint is given by $\xi = [v^T \ \mathbf{0}]^T$. For a rotation by an angle θ about the axis we get a transformation of

$$\hat{\xi} = \begin{bmatrix} \mathbf{0} & v \\ \mathbf{0} & 0 \end{bmatrix}, \qquad e^{\hat{\xi}\theta} = \begin{bmatrix} I & v\theta \\ \mathbf{0} & 1 \end{bmatrix}$$

11.3.2 Forward Kinematics

The full transformation from the end (or *tool*) frame to the base (or *stationary*) frame for a manipulator chain can be computed using the *product of exponentials* formula as follows.

- **Initial configuration:**

  We start by laying out the manipulator in a base configuration (where all joint angles $\theta_i = 0$). We then define the initial transformation from the tool frame to the stationary frame

$$g_{\text{ST}}(0) = \begin{bmatrix} R_{\text{ST}(0)} & p_{\text{ST}}(0) \\ \mathbf{0} & 1 \end{bmatrix}$$

- **Joint Motion Application:**

  The motion of each joint is then applied to the manipulator starting with the joints **closest** to the tool frame and work backwards toward the stationary frame. This is to allow for the fact that joints only affect the links further along the manipulator chain. Constructing the full position of the manipulator in this way is illustrated in the figure below.

Initial Configuration:
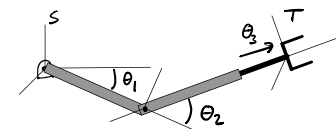
$$g_{ST}(\theta) = g_{ST}(0)$$

Joint 3: (prismatic)

$$g_{ST}(\theta) = e^{\hat{\xi}_3 \theta_3} g_{ST}(0)$$

JOINT 2:

$$g_{ST}(\theta) = e^{\hat{\xi}_2 \theta_2} e^{\hat{\xi}_3 \theta_3} g_{ST}(0)$$

JOINT 1:

$$g_{ST}(\theta) = e^{\hat{\xi}_1 \theta_1} e^{\hat{\xi}_2 \theta_2} e^{\hat{\xi}_3 \theta_3} g_{ST}(0)$$

- **Product of Exponentials:**

  The full coordinate transformation from the tool to the stationary frame is then given by the product of exponentials formula

  $$g_{\text{ST}}(\theta) = e^{\hat{\xi}_1 \theta_1} \cdots e^{\hat{\xi}_n \theta_n} g_{\text{ST}}(0)$$

  To make the application of this formula more clear we give two examples, the SCARA manipulator and a standard elbow manipulator.

1. **SCARA manipulator**

Figure 3.3: SCARA manipulator in its reference configuration.

We compute the kinematics for $l_0 = 1$, $l_1 = 1$, and $l_2 = 1$.

$$\omega_1 = \omega_2 = \omega_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \hat{\omega}_i = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$q_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 0 \\ l_1 = 1 \\ 0 \end{bmatrix}, \quad q_3 = \begin{bmatrix} 0 \\ l_1 + l_2 = 2 \\ 0 \end{bmatrix},$$

Revolute (rotational) joint: for $i = 1, 2, 3$

$$\hat{\xi}_i = \begin{bmatrix} \hat{\omega}_i & -\hat{\omega}_i q_i \\ \mathbf{0} & 0 \end{bmatrix}$$

$$\hat{\xi}_1 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{\xi}_2 = \begin{bmatrix} 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \hat{\xi}_3 = \begin{bmatrix} 0 & -1 & 0 & 2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Prismatic (linear) Joint:

$$\hat{\xi}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Initial Configuration:

$$g_{ST}(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & l_1 + l_2 \\ 0 & 0 & 1 & l_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Product of Exponentials:

$$e^{\hat{\xi}_1 \theta_1} e^{\hat{\xi}_2 \theta_2} e^{\hat{\xi}_3 \theta_3} e^{\hat{\xi}_4 \theta_4} = \begin{bmatrix} e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} e^{\hat{\omega}_3 \theta_3} & q \\ \mathbf{0} & 1 \end{bmatrix}$$

where

$$e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} e^{\hat{\omega}_3 \theta_3} = \begin{bmatrix} c(\theta_1 + \theta_2 + \theta_3) & -s(\theta_1 + \theta_2 + \theta_3) & 0 \\ s(\theta_1 + \theta_2 + \theta_3) & c(\theta_1 + \theta_2 + \theta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$
\begin{aligned}
q &= e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} e^{\hat{\omega}_3 \theta_3} v_4 \theta_4 \\
&\quad - e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} \left(I - e^{\hat{\omega}_3 \theta_3} \right) \hat{\omega}_3^2 q_3 \\
&\quad - e^{\hat{\omega}_1 \theta_1} \left(I - e^{\hat{\omega}_2 \theta_2} \right) \hat{\omega}_2^2 q_2 - \left(I - e^{\hat{\omega}_1 \theta_1} \right) \hat{\omega}_1^2 q_1 \\
&= e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} e^{\hat{\omega}_3 \theta_3} \left(v_4 \theta_4 + \hat{\omega}_3^2 q_3 \right) \\
&\quad - e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} \left(\hat{\omega}_3^2 q_3 - \hat{\omega}_2^2 q_2 \right) - e^{\hat{\omega}_1 \theta_1} \left(\hat{\omega}_2^2 q_2 - \hat{\omega}_1^2 q_1 \right) - \hat{\omega}_1^2 q_1 \\
&= \begin{bmatrix} c(\theta_1 + \theta_2 + \theta_3) & -s(\theta_1 + \theta_2 + \theta_3) & 0 \\ s(\theta_1 + \theta_2 + \theta_3) & c(\theta_1 + \theta_2 + \theta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \left(v_4 \theta_4 + \hat{\omega}_3^2 q_3 \right) \\
&\quad - \begin{bmatrix} c(\theta_1 + \theta_2) & -s(\theta_1 + \theta_2) & 0 \\ s(\theta_1 + \theta_2) & .c(\theta_1 + \theta_2) & 0 \\ 0 & 0 & 1 \end{bmatrix} \left(\hat{\omega}_3^2 q_3 - \hat{\omega}_2^2 q_2 \right) \\
&\quad . - \begin{bmatrix} c(\theta_1) & -s(\theta_1) & 0 \\ s(\theta_1) & .c(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \left(\hat{\omega}_2^2 q_2 - \hat{\omega}_1^2 q_1 \right) - \hat{\omega}_1^2 q_1 \\
&= \begin{bmatrix} c(\theta_1 + \theta_2 + \theta_3) & -s(\theta_1 + \theta_2 + \theta_3) & 0 \\ s(\theta_1 + \theta_2 + \theta_3) & c(\theta_1 + \theta_2 + \theta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \\ \theta_4 \end{bmatrix} \\
&\quad + \begin{bmatrix} c(\theta_1 + \theta_2) & -s(\theta_1 + \theta_2) & 0 \\ s(\theta_1 + \theta_2) & .c(\theta_1 + \theta_2) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} c(\theta_1) & -s(\theta_1) & 0 \\ s(\theta_1) & .c(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 2s(\theta_1 + \theta_2 + \theta_3) - s(\theta_1 + \theta_2) - s(\theta_1) \\ -2c(\theta_1 + \theta_2 + \theta_3) + c(\theta_1 + \theta_2) + c(\theta_1) \\ \theta_4 \end{bmatrix}
\end{aligned}
$$

Forward Kinematics: Let $\theta = [\theta_1 \; \theta_2 \; \theta_3]^T$

$$g_{ST}(\theta) = e^{\hat{\xi}_1\theta_1}e^{\hat{\xi}_2\theta_2}e^{\hat{\xi}_3\theta_3}e^{\hat{\xi}_4\theta_4}g_{ST}(0)$$

$$= \begin{bmatrix} c(\mathbf{1}^T\theta) & -s(\mathbf{1}^T\theta) & 0 & 2s(\mathbf{1}^T\theta) - s(\theta_1+\theta_2) - s(\theta_1) \\ s(\mathbf{1}^T\theta) & c(\mathbf{1}^T\theta) & 0 & -2c(\mathbf{1}^T\theta) + c(\theta_1+\theta_2) + c(\theta_1) \\ 0 & 0 & 1 & \theta_4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c(\mathbf{1}^T\theta) & -s(\mathbf{1}^T\theta) & 0 & -s(\theta_1+\theta_2) - s(\theta_1) \\ s(\mathbf{1}^T\theta) & c(\mathbf{1}^T\theta) & 0 & c(\theta_1+\theta_2) + c(\theta_1) \\ 0 & 0 & 1 & 1+\theta_4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2. **Elbow Manipulator**

Figure 3.4: Elbow manipulator.

We compute the kinematics for $l_0 = 1$, $l_1 = 1$, and $l_2 = 1$.

$$\hat{\omega}_1 = \hat{\omega}_4 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \hat{\omega}_6 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \qquad \hat{\omega}_2 = \hat{\omega}_3 = \hat{\omega}_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix},$$

$$q_1 = \begin{bmatrix} 0 \\ 0 \\ l_0 = 1 \end{bmatrix}, \qquad q_2 = \begin{bmatrix} 0 \\ l_1 = 1 \\ l_0 = 1 \end{bmatrix}, \qquad q_3 = \begin{bmatrix} 0 \\ l_1 + l_2 = 2 \\ l_0 = 1 \end{bmatrix},$$

Revolute (rotational) joint: for $i = 1, 2, 3, 4, 5, 6$

$$\hat{\xi}_1 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{\xi}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \hat{\xi}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{\xi}_4 = \begin{bmatrix} 0 & -1 & 0 & 2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{\xi}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \hat{\xi}_6 = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Initial Configuration:

$$g_{ST}(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & l_1 + l_2 \\ 0 & 0 & 1 & l_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Product of Exponentials:

$$e^{\hat{\xi}_1 \theta_1} e^{\hat{\xi}_2 \theta_2} e^{\hat{\xi}_3 \theta_3} e^{\hat{\xi}_4 \theta_4} e^{\hat{\xi}_5 \theta_5} e^{\hat{\xi}_6 \theta_6} = \begin{bmatrix} R & q \\ \mathbf{0} & 1 \end{bmatrix}$$
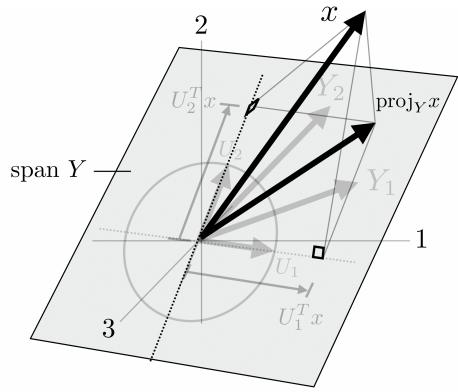
where

$$R = e^{\hat{\omega}_1 \theta_1} e^{\hat{\omega}_2 \theta_2} e^{\hat{\omega}_3 \theta_3} e^{\hat{\omega}_4 \theta_4} e^{\hat{\omega}_5 \theta_5} e^{\hat{\omega}_6 \theta_6}$$
$$= R^{123} R^{45} R^6$$

$$q = e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2}e^{\hat{\omega}_3\theta_3}e^{\hat{\omega}_4\theta_4}e^{\hat{\omega}_5\theta_5}\left(-\hat{\omega}_6^2 q_6 + \hat{\omega}_5^2 q_5\right)$$

$$+ e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2}e^{\hat{\omega}_3\theta_3}e^{\hat{\omega}_4\theta_4}\left(-\hat{\omega}_5^2 q_5 + \hat{\omega}_4^2 q_4\right)$$

$$+ e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2}e^{\hat{\omega}_3\theta_3}\left(-\hat{\omega}_4^2 q_4 + \hat{\omega}_3^2 q_3\right)$$

$$+ e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2}\left(-\hat{\omega}_3^2 q_3 + \hat{\omega}_2^2 q_2\right)$$

$$+ e^{\hat{\omega}_1\theta_1}\left(-\hat{\omega}_2^2 q_2 + \hat{\omega}_1^2 q_1\right)$$

$$+ \left(-\hat{\omega}_1^2 q_1\right)$$

$$= R^{123}R^{45}\begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} + R^{123}R^4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + R^{123}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + R^{12}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + R^1\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= R^{123}R^{45}\begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} -s\theta_1 c(\theta_2+\theta_3) - s\theta_1 c\theta_2 \\ c\theta_1 c(\theta_2+\theta_3) + c\theta_1 c\theta_2 \\ -s(\theta_2+\theta_3) - s\theta_2 + 1 \end{bmatrix}$$

$$R^1 = e^{\hat{\omega}_1\theta_1} = \begin{bmatrix} c\theta_1 & -s\theta_1 & 0 \\ s\theta_1 & c\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R^{12} = e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2} = \begin{bmatrix} c\theta_1 & -s\theta_1 c\theta_2 & -s\theta_1 s\theta_2 \\ s\theta_1 & c\theta_1 c\theta_2 & c\theta_1 s\theta_2 \\ 0 & -s\theta_2 & c\theta_2 \end{bmatrix}$$

$$R^{123} = e^{\hat{\omega}_1\theta_1}e^{\hat{\omega}_2\theta_2}e^{\hat{\omega}_3\theta_3} = \begin{bmatrix} c\theta_1 & -s\theta_1 c(\theta_2+\theta_3) & -s\theta_1 s(\theta_2+\theta_3) \\ s\theta_1 & c\theta_1 c(\theta_2+\theta_3) & c\theta_1 s(\theta_2+\theta_3) \\ 0 & -s(\theta_2+\theta_3) & c(\theta_2+\theta_3) \end{bmatrix}$$

$$R^4 = e^{\hat{\omega}_4\theta_4} = \begin{bmatrix} c\theta_4 & -s\theta_4 & 0 \\ s\theta_4 & c\theta_4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R^{45} = e^{\hat{\omega}_4\theta_4}e^{\hat{\omega}_5\theta_5} = \begin{bmatrix} c\theta_4 & -s\theta_4 c\theta_5 & -s\theta_4 s\theta_5 \\ s\theta_4 & c\theta_4 c\theta_5 & c\theta_4 s\theta_5 \\ 0 & -s\theta_5 & c\theta_5 \end{bmatrix}$$

$$R^6 = e^{\hat{\omega}_6\theta_6} = \begin{bmatrix} c\theta_6 & 0 & s\theta_6 \\ 0 & 0 & 0 \\ -s\theta_6 & 0 & c\theta_6 \end{bmatrix}$$

Forward Kinematics:

$$g_{ST}(\theta) = e^{\hat{\xi}_1\theta_1}e^{\hat{\xi}_2\theta_2}e^{\hat{\xi}_3\theta_3}e^{\hat{\xi}_4\theta_4}e^{\hat{\xi}_5\theta_5}e^{\hat{\xi}_6\theta_6}g_{ST}(0)$$

$$= \begin{bmatrix} R & q \\ \mathbf{0} & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} R^{123}R^{45}R^6 & R^{123}R^{45}\left(R^6\begin{bmatrix}0\\2\\1\end{bmatrix} + \begin{bmatrix}0\\-2\\0\end{bmatrix}\right) + \begin{bmatrix}-s\theta_1 c(\theta_2+\theta_3) - s\theta_1 c\theta_2 \\ c\theta_1 c(\theta_2+\theta_3) + c\theta_1 c\theta_2 \\ -s(\theta_2+\theta_3) - s\theta_2 + 1\end{bmatrix} \\ \begin{bmatrix}0 & 0 & 0\end{bmatrix} & 1 \end{bmatrix}$$
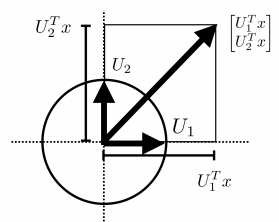
Define subspace
$$\operatorname{span}\begin{bmatrix} | & | \\ Y_1 & Y_2 \\ | & | \end{bmatrix}$$

ortho-normalize basis
$$\underbrace{\begin{bmatrix} | & | \\ U_1 & U_2 \\ | & | \end{bmatrix}}_{\substack{\text{orthonormal} \\ \text{basis}}} = \begin{bmatrix} | & | \\ Y_1 & Y_2 \\ | & | \end{bmatrix}\begin{bmatrix} Y^T Y \end{bmatrix}^{-\frac{1}{2}}$$

project onto lower dim space
$$\begin{bmatrix} - & U_1^T & - \\ - & U_2^T & - \end{bmatrix}\begin{bmatrix} | \\ x \\ | \end{bmatrix} = \begin{bmatrix} U_1^T x \\ U_2^T x \end{bmatrix}$$

return to original space
$$\operatorname{proj}_Y x = \begin{bmatrix} | & | \\ U_1 & U_2 \\ | & | \end{bmatrix}\begin{bmatrix} - & U_1^T & - \\ - & U_2^T & - \end{bmatrix}\begin{bmatrix} | \\ x \\ | \end{bmatrix}$$
$$= \begin{bmatrix} | & | \\ Y_1 & Y_2 \\ | & | \end{bmatrix}\begin{bmatrix} Y^T Y \end{bmatrix}^{-1}\begin{bmatrix} - & Y_1^T & - \\ - & Y_2^T & - \end{bmatrix}\begin{bmatrix} | \\ x \\ | \end{bmatrix}$$
$$= Y(Y^T Y)^{-1} Y^T x$$