Imperial College London

Mathematics: Year 2

Linear Algebra

Xin Wang

December 5, 2020

Abstract

As mentioned in the previous chapter, if no solution is present i.e. the system is inconsistent, then it has to be dealt with another way. No exact solutions can be found, it has to be approximated.

Contents

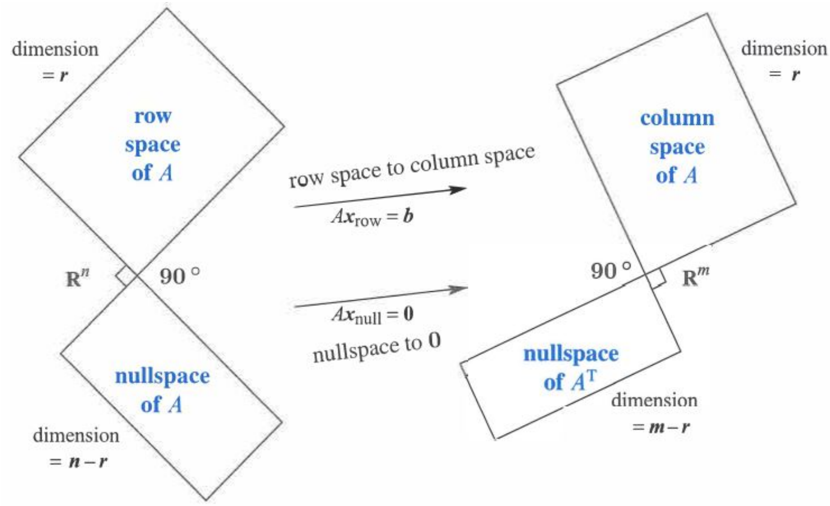
1 Orthogonality

This concept is moved to orthogonal subspaces and orthogonal bases and orthogonal matrices. Recall that the two vectors are orthogonal when their dot product is zero: $v.w = v^T w = 0$. Since it is perpendicular, Pythagoras can be used:

$$||v||^2 + ||w||^2 = ||v + w||^2$$

Recall the four fundamental subspaces that reveal what a matrix really does. A matrix multiplies a vector: A times x.

- At the first level this is only numbers.
- At the second level Ax is a combination of column vectors.
- At third level is with subspaces.



With the concept of orthogonality:

- **The row space is perpendicular to the nullspace**.

  Every row of A is perpendicular to every solution of $Ax = 0$. That gives the 90 degree angle on the left side of the figure. This perpendicularity of subspaces is what is of interest.

- **The column space is perpendicular to the nullspace of** A^T.

  When b is outside the column space, solving $Ax = b$ is impossible. This nullspace of A^T comes into its own. It will contain the error $e = b - Ax$ in the "least-squares" solution. Least squares is the key application of linear algebra in this chapter.

Orthogonal subspaces: $v^T w = 0$ for all v in V and all w in W.

Note: Orthogonality is impossible when $dim(V)+dim(W) > dim(wholespace)$. Two planes (dimensions 2 and 2 in R^3) cannot be orthogonal subspaces.

The crucial examples for linear algebra come fr om the four fundamental subspaces. Zero is the only point where the nullspace meets the row space.

Every vector x in the nullspace is perpendicular to every row of A, because $Ax = 0$. The null space $N(A)$ and the row space $C(A^T)$ are orthogonal subspaces of R^n.

$$Ax = \begin{bmatrix} \textbf{row 1} \\ \vdots \\ \textbf{row } m \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \longleftarrow \quad (\textbf{row 1}) \cdot x \text{ is zero} \\ \\ \longleftarrow \quad (\textbf{row } m) \cdot x \text{ is zero} \end{array}$$

1.1 Orthogonal complements

The fundamental subspaces are more than just orthogonal (in pairs). Their dimensions are also just right.

Consider two lines that are perpendicular in R^3. Those lines **could not be** the row space and nullspace of a 3×3 matrix. The lines have dimensions 1 and 1, adding to 2. But the correct dimensions $n - r$ must add to $n = 3$.

The fundamental subspaces of a 3 by 3 matrix have dimensions 2 and 1, or 3 and 0. Those pairs of subspaces are not only orthogonal, they are orthogonal complements.

Orthogonal complement (of a subspace V): Denoted $V^{\perp}$, contains every vector that is perpendicular to V.

***$N(A)$ is the orthogonal complement of the row space $C(A^{\mathrm{T}})$* (in $\mathbf{R}^n$).**

***$N(A^{\mathrm{T}})$ is the orthogonal complement of the column space $C(A)$* (in $\mathbf{R}^m$)**

The two subspaces are orthogonal complements if:

$\dim N(A) + \dim C(AT) = \dim R^n = n$ and $\dim N(A^T) + \dim C(A) = \dim R^m = m$.

Example 1: Show that nullspace and row space of $A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$ are orthogonal subspaces of $\Re^3$.

- Determine the rank: Rank = 1.
- Find the dimensions of the column: $\dim C(A^T) = 1$
- Find the dimensions of the nullspace: $\dim N(A) = n - r = 3 - 1 = 2$
- Find if every vector in this nullspace plane is orthogonal to the vectors in column space, then the plane is orthogonal:
  - Use elimination processes on A:

$$A \sim \begin{bmatrix} 1 & 2 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

  - Find special solutions since x_2 and x_3 are free variables:

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}$$

  - Multiple with column space vector to check if it is orthogonal or not:

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = 0 \text{ and } \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = 0$$

2 Matrix A^TA and projection

As mentioned earlier with the problem of $Ax = b$ when there is no solution. To "solve" the equation, the best possible approximation has to be found. These problems occur often e.g. measurements of the position of a satellite: large m. These gives m equations to estimate the n parameters determining the trajectory, where n is much smaller i.e. *tall matrix*.

In order to approximate the solutions, several concepts need to be established.

2.1 A^TA and basis

Recall that **basis** is linearly independent vectors that span the space.

It is already known that the matrix A^TA is square $[n \times n]$ and symmetric.

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}}$$
$$A^TA\hat{\underline{\mathbf{x}}} = A^T\underline{\mathbf{b}}$$

where $\hat{\underline{\mathbf{x}}}$ is an approximation of $\underline{\mathbf{x}}$. The approximation is such that the error in the approximation is as small as possible.

The area of interest is whether if the matrix is invertible? If not, what is it the nullspace?

Consider the two examples:

Case A:

$$A^TA = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 8 \\ 8 & 30 \end{pmatrix}$$

Here the matrix A^TA is a full column rank, the columns are independent. A it is invertible.

Case B:

$$A^TA = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 12 \end{pmatrix}$$

Here the matrix A^TA is a not a full column rank since Column 2 is a multiple of Column 1, meaning that A it is not invertible.

One property implies the other :

- Any n independent vectors in R^n must span R^n. So the columns are a basis.
- Any n vectors that span R^n must be independent. So the columns are a basis.

Considering the vector columns shown in the two cases:

- If the n columns of A are independent, they span R^n . So $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ is solvable i.e. the solution x is unique.
- If the n columns span R^n , the columns are independent. So $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has only one solution.

In summary, uniqueness implies existence and existence implies uniqueness. Because there are n pivots and no free variables, the nullspace contains only $x = 0$ i.e. an invertible matrix has $N(A) = \underline{\mathbf{0}}$.

$$N(A^T A) = N(A) \text{ and } rank(A^T A) = rank(A)$$

If A has independent columns, then the only solution to $A^T A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ is $\underline{\mathbf{x}} = \underline{\mathbf{0}}$.

2.2 Projection

To visualise projection, consider the projections of $b = (2, 3, 4)$ onto the z axis and the xy plane?

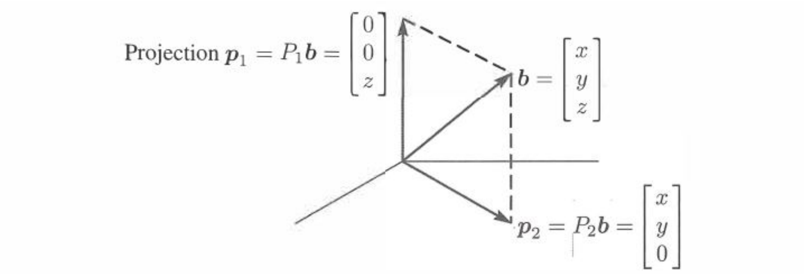


Figure 4.5: The projections $\boldsymbol{p}_1 = P_1\boldsymbol{b}$ and $\boldsymbol{p}_2 = P_2\boldsymbol{b}$ onto the z axis and the xy plane.

- The projection onto the z axis denoted p_1:

$$\boldsymbol{p}_1 = P_1\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{z} \end{bmatrix}$$

- The projection onto the xy axis denoted p_2:

$$\boldsymbol{p}_2 = P_2\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}$$

where the **projection matrix** extracts the wanted components from b.

Note that the two projected lines p_1 and p_2 are orthogonal complements. Their dimensions add to $1+2 = 3$. Every vector b in the whole space is the sum of its parts in the two subspaces, for example the projections p_1 and p_2 are exactly those two parts of b.

This example shows the objective: find the part p in each subspace, and the projection matrix P that produces that part $p = Pb$.

Every subspace of R^m has its own $[m \times m]$ projection matrix. To compute P, a good description of the subspace that it projects onto is needed. The best description of a subspace is a basis.

2.2.1 Projection onto 1D

The key to projection is orthogonality and it well known that the perpendicular distance is the shortest distance between two lines. In estimation, it is of interest the point given by $\underline{\mathbf{p}}$ i.e. the projection of $\underline{\mathbf{b}}$ onto $\underline{\mathbf{a}}$. **The line from b to p is perpendicular to a**

The particular reason is that this is the point on $\underline{\mathbf{a}}$ nearest to $\underline{\mathbf{b}}$. Recall the usual objective is to obtain $\underline{\mathbf{b}}$, but this is not available, thus to points on $\underline{\mathbf{a}}$, the nearest point on $\underline{\mathbf{a}}$ is the projection $\underline{\mathbf{p}}$.

The projection $\underline{\mathbf{p}}$ will be some multiple of $\underline{\mathbf{a}}$:

$$\underline{\mathbf{p}} = \underline{\hat{\mathbf{x}}} \times \underline{\mathbf{a}}$$

These three steps will lead to all projection matrices:

- Find $\underline{\hat{\mathbf{x}}}$

$$\underline{\hat{\mathbf{x}}} = \frac{\underline{\mathbf{a}}.\underline{\mathbf{b}}}{\underline{\mathbf{a}}.\underline{\mathbf{a}}} = \frac{\underline{\mathbf{a}}^T\underline{\mathbf{b}}}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}$$

- Computing $\underline{\hat{\mathbf{x}}}$ will give the vector $\underline{\mathbf{p}}$

$$\underline{\mathbf{p}} = \frac{\underline{\mathbf{a}}^T\underline{\mathbf{b}}}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}\underline{\mathbf{a}} = \frac{\underline{\mathbf{a}}\underline{\mathbf{a}}^T}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}\underline{\mathbf{b}}$$

  where $(\underline{\mathbf{a}}^T\underline{\mathbf{b}})\underline{\mathbf{a}} = \underline{\mathbf{a}}(\underline{\mathbf{a}}^T\underline{\mathbf{b}})$

- From the formula for $\underline{\mathbf{p}}$, the projection matrix P can be found

$$P = \frac{\underline{\mathbf{a}}\underline{\mathbf{a}}^T}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}$$

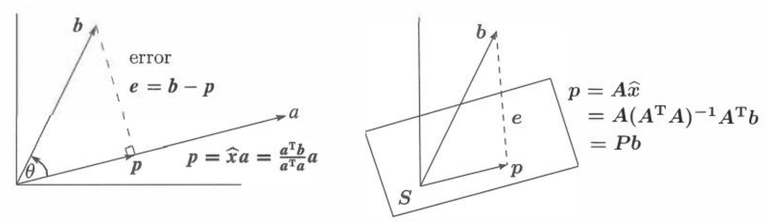


Figure 1: The projection p of b onto a line and onto S = column space of A

Note that if projection occurs twice onto $\underline{\mathbf{a}}$? The first time from $\underline{\mathbf{b}}$ to $\underline{\mathbf{p}}$ but the second time - nothing happens. This is due to that there is a line through $\underline{\mathbf{a}}$ already.

$$P^2\underline{\mathbf{b}} = P(P\underline{\mathbf{b}}) = P\underline{\mathbf{p}} = \underline{\mathbf{p}} = P\underline{\mathbf{b}}$$

The difference real and approximation becomes the resulting error $\underline{\mathbf{e}} \neq 0$.

$$\underline{\mathbf{e}} = \underline{\mathbf{b}} - \underline{\hat{\mathbf{x}}\mathbf{a}}$$

2.2.2 Projection onto 2D

Returning to the problem $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ which has no solution. To find the closest thing to a solution:

- Consider that $A\underline{\mathbf{x}}$ is in the column space of $C(A)$, and look for $\underline{\mathbf{p}}$ which is the closest vector in $C(A)$ to $\underline{\mathbf{b}}$
- Solve $A\underline{\hat{\mathbf{x}}} = \underline{\mathbf{p}}$ where $\underline{\mathbf{p}}$ is the projection of $\underline{\mathbf{b}}$ onto the column space of A.

$$A^T A\underline{\hat{\mathbf{x}}} = A^T\underline{\mathbf{b}}$$

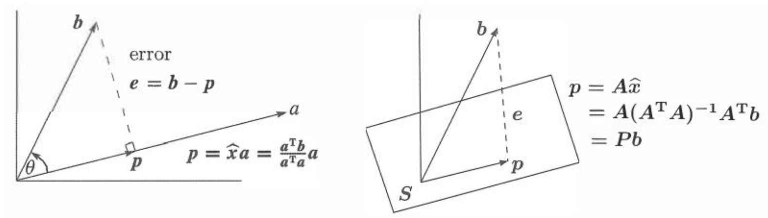


Figure 2: The projection p of b onto a line and onto S = column space of A

The steps is the same as the previous topic. These three steps will lead to all projection matrices:

- Find $\underline{\hat{\mathbf{x}}}$

$$\underline{\hat{\mathbf{x}}} = (A^T A)^{-1} A^T\underline{\mathbf{b}} = \frac{\underline{\mathbf{a}}.\underline{\mathbf{b}}}{\underline{\mathbf{a}}.\underline{\mathbf{a}}} = \frac{\underline{\mathbf{a}}^T\underline{\mathbf{b}}}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}$$

- Computing $\underline{\hat{\mathbf{x}}}$ will give the vector $\underline{\mathbf{p}}$

$$\underline{\mathbf{p}} = A\underline{\hat{\mathbf{x}}} = A(A^T A)^{-1} A^T\underline{\mathbf{b}} = \frac{\underline{\mathbf{a}}^T\underline{\mathbf{b}}}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}\underline{\mathbf{a}} = \frac{\underline{\mathbf{a}\mathbf{a}}^T}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}\underline{\mathbf{b}}$$

  where $(\underline{\mathbf{a}}^T\underline{\mathbf{b}})\underline{\mathbf{a}} = \underline{\mathbf{a}}(\underline{\mathbf{a}}^T\underline{\mathbf{b}})$

- From the formula for $\underline{\mathbf{p}}$, the projection matrix P can be found

$$P = A(A^T A)^{-1} A^T = \frac{\underline{\mathbf{a}\mathbf{a}}^T}{\underline{\mathbf{a}}^T\underline{\mathbf{a}}}$$

Note that the projection matrix had two properties:

- P is symmetric
- $P^2 = P$

3 Least-squares method or Linear Regression

It often happens that $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has no solution. The usual reason is: too many equations. The matrix A has more rows than columns. There are more equations than unknowns i.e m is greater than n.

When $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has no solution, multiply by A^T and solve $A^T A\underline{\mathbf{x}} = A^T\underline{\mathbf{b}}$. A crucial application of least squares is fitting a straight line to m points.

Worked Example: Find the closest line to the points $(0, 6)$, $(1, 0)$, and $(2, 0)$.

No straight line $b = C + Dt$ goes through those three points. We are asking for two numbers C and D that satisfy three equations from the three coordinates listed.

Here are the three equations at $t = 0,\ 1,\ 2$ to match the given values $\underline{\mathbf{b}} = 6, 0, 0$:

$$\begin{aligned} C + D.0 &= 6 \\ C + D.1 &= 0 \\ C + D.2 &= 0 \end{aligned}$$

Notice that this 3 by 2 system has no solution: $\underline{\mathbf{b}} = 6, 0, 0$ is not a combination of the columns $(1, 1, 1)$ and $(0, 1, 2)$:

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}} \Rightarrow \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$$

To find the best fit solution, solve $A^T A\underline{\hat{\mathbf{x}}} = A^T\underline{\mathbf{b}}$: where $\underline{\hat{\mathbf{x}}} = \begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix}$ will be the "best" approximations to the impossible $\begin{pmatrix} C \\ D \end{pmatrix}$.

$$\begin{aligned} A^T A\underline{\hat{\mathbf{x}}} = \underline{\mathbf{b}} \Rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \underline{\hat{\mathbf{x}}} &= \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} \\ = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix} &= \begin{pmatrix} 6 \\ 0 \end{pmatrix} \end{aligned}$$

Solve the two equations to find the values of $\hat{C}$ and $\hat{D}$.

$$\begin{aligned} 3\hat{C} + 3\hat{D} &= 6 \\ 3\hat{C} + 5\hat{D} &= 0 \end{aligned}$$

Thus:

$$\underline{\hat{\mathbf{x}}} = \begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

So the though least square approximation, the points closest to $\underline{\mathbf{b}}$ denoted $\hat{\underline{\mathbf{b}}}$ is found:

$$
\begin{aligned}
A\hat{\underline{\mathbf{x}}} = \hat{\underline{\mathbf{b}}} \Rightarrow \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} \\
= \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}
\end{aligned}
$$

3.1 Minimising error

How do we make the error $\underline{\mathbf{e}} = \underline{\mathbf{b}} - A\hat{\underline{\mathbf{x}}}$ as small as possible? Recall that the point nearest to b is basically the projection p.

The smallest possible error is perpendicular to the columns defined as:

$$\underline{\mathbf{e}} = \underline{\mathbf{b}} - \underline{\mathbf{p}}$$

4 Orthogonality and QR

There are two objective to this section:

- The first is to see why orthogonality is good.
- The second goal is to construct the orthogonal vectors.

Dot products are zero, so $A^T A$ will be diagonal. It becomes so easy to find $\underline{\hat{\mathbf{x}}}$ and $\underline{\mathbf{p}} = A\underline{\hat{\mathbf{x}}}$.

Orthogonal vectors are constructed from the Gram-Schmidt process which chooses combinations of the original basis vectors to produce right angles.

4.1 Orthonormal vectors and orthogonal matrices

Recall that a basis consists of independent vectors that span the space. The basis vectors could meet at any angle except at 0 and 180.

The vectors $q_1, \dots, q_n$ are orthogonal when their dot products $q_i.q_j$ are zero i.e. $q_i^T q_j = 0$ whenever $i \neq j$. The **orthogonal unit vectors** are found by dividing each vector by its length.

Orthonomality: A matrix with orthonormal columns is assigned the special letter Q.

$$\boldsymbol{q}_i^{\mathrm{T}} \boldsymbol{q}_j = \begin{cases} 0 & \text{when } i \neq j \quad \text{(orthogonal vectors)} \\ 1 & \text{when } i = j \quad \text{(unit vectors: } \|\boldsymbol{q}_i\| = 1) \end{cases}$$

The reason for Q is because it is easy to work with since $Q^T Q = I$. A matrix Q with orthonormal columns satisfies $Q^T Q = I$:

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{bmatrix} -\boldsymbol{q}_1^{\mathrm{T}}- \\ -\boldsymbol{q}_2^{\mathrm{T}}- \\ -\boldsymbol{q}_n^{\mathrm{T}}- \end{bmatrix} \begin{bmatrix} | & | & \mathbf{1} \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_n \\ | & | & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

When Q is square, $Q^T Q = I$ means that $Q^T = Q^{-1}$ i.e. transpose = inverse. In this case, Q is an **orthogonal matrix**.

Example 1: Check if the given matrix $Q = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ is orthogonal or not.

Matrix Q is almost orthogonal, it needs to be normalised first:

$$\hat{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

Normalised matrix Q satisfies the orthogonality property:

$$Q \times Q^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \times \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that rotations preserve the length of every vector. So do reflections. So do permutations. So does multiplication by any orthogonal matrix Q i.e. lengths and angles don't change.

If Q has orthonormal columns ($Q^TQ = I$), it leaves lengths unchanged:

$$||Q\mathbf{x}|| = ||\mathbf{x}||$$

This means that Q also preserves dot products:

$$(Q\mathbf{x})^T(Q\mathbf{y}) = \mathbf{x}^TQ^TQ\mathbf{y} = \mathbf{x}^T\mathbf{y}$$

thus simplifying to

$$Q^TQ = I$$

4.1.1 Projections Using Orthonormal Bases: Q replaces A

Recall:

- How a projection works: we would like to project onto the column space of a matrix, as we did, for example, in the least-squares algorithm.

- A system of linear equations $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has no solution, but we can find the nearest point to $\underline{\mathbf{b}}$ in the column space of A, so that $A^TA\hat{\underline{\mathbf{x}}} = A^T\underline{\mathbf{b}}$ does have a solution.

Suppose the basis vectors are actually orthonormal. Then A^TA simplifies to $Q^TQ = I$. This means that the least squares solution of $Qx = b$ is $x = Q^Tb$ and the projection matrix is QQ^T.

There are no matrices to invert. This is the point of an orthonormal basis. The best $\underline{\mathbf{x}} = Q^T\underline{\mathbf{b}}$ just has dot products of $q_1, \ldots, q_n$ with $\underline{\mathbf{b}}$. The result is 1-dimensional projections.

4.2 Gram-Schmidt process

The immense benefit of using Q has already been stated and all these relies on the fact that the **basis are orthonormal**. The Gram-Schmidt process is a way to create these orthonormal vectors.

Start with three **independent** vectors a, b, c. We intend to construct three orthogonal vectors A, B, C. Then we divide A, B, C by their lengths, this produces three orthonormal vectors $q_1 = \frac{A}{||A||}$, $q_2 = \frac{b}{||b||}$, $q_3 = \frac{C}{||C||}$.

Begin by choosing $A = \mathbf{a}$. This first direction does not really matter and can be accepted as it comes. The next direction B **must be perpendicular to** A. Start with $\mathbf{b}$ and subtract its projection along A - leaving the perpendicular part, which is the orthogonal vector B:

$$B = \mathbf{b} - \frac{A^T\mathbf{b}}{A^TA}A \tag{1}$$

A and B are orthogonal now.

The third direction starts with c. This is not a combination of A and B (because c is not a combination of a and b). c is not perpendicular to A and B so subtract off its components in those two directions to get a perpendicular direction C:

$$C = \mathbf{c} - \frac{A^T\mathbf{c}}{A^TA}A - \frac{B^T\mathbf{c}}{B^TB}B \tag{2}$$

The Gram-Schmidt process can be summarised as: *Subtract from every new vector its projections in the directions already set.*

At the end, or immediately when each one is found, divide the orthogonal vectors A, B, C by their lengths. The resulting vectors q_1, q_2, q_3 are orthonormal.

Example 1: Consider the basis of $\Re^3$

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Obtain an orthonormal basis by implementing the Gram-Schmidt process.

1. A:

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

2. B:

$$B = \mathbf{b} - \frac{A^T\mathbf{b}}{A^TA}A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{5}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$$

Note that both $\frac{1}{5}\begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$ are clearly orthogonal so B can be simplified to $B = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$.

3. C:

$$C = \mathbf{c} - \frac{A^T\mathbf{c}}{A^TA}A - \frac{B^T\mathbf{c}}{B^TB}B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{15}{30}\begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} - \frac{5}{5}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

To get an orthonormal basis, divide each vector by its length:

$$A = \tfrac{1}{\|\underline{\mathbf{x}}_1\|}\underline{\mathbf{x}}_1 = \tfrac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

$$B = \tfrac{1}{\|\underline{\mathbf{x}}_2\|}\underline{\mathbf{x}}_2 = \tfrac{1}{\sqrt{30}}\begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$$

$$C = \tfrac{1}{\|\underline{\mathbf{x}}_3\|}\underline{\mathbf{x}}_3 = \tfrac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

4.3 Factorization $A = QR$

We started with a matrix A, whose columns were $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. Through the Gram-Schmidt process, a matrix Q is formed from the independent vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$. This hints that there must be a relationship between the two matrices.

Given the relations between the columns of A and Q, we can write these as a matrix product:

$$A = QR$$

Note the key points of Gram-Schmidt:

- The vectors $\mathbf{a}$ and A and $\mathbf{q}_1$ are all along a single line. This means that $\mathbf{q}_1$ is a linear combination of $\mathbf{a}$.
- The vectors $\mathbf{a}$, $\mathbf{b}$ and A, B and $\mathbf{q}_1$, $\mathbf{q}_2$ are all in the same plane. This means that $\mathbf{q}_2$ is a linear combination of $\mathbf{a}$ and $\mathbf{b}$.
- The vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and A, B, C and $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ are in one subspace (dimension 3). Note that $\mathbf{q}_3$ is orthogonal to $\mathbf{q}_1$ and $\mathbf{q}_2$ thus also orthogonal to $\mathbf{a}$ and $\mathbf{b}$.

$A = QR$ is Gram-Schmidt in a nutshell:

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}} a & q_1^{\mathrm{T}} b & q_1^{\mathrm{T}} c \\ & q_2^{\mathrm{T}} b & q_2^{\mathrm{T}} c \\ & & q_3^{\mathrm{T}} c \end{bmatrix}$$

Multiply by Q^T to recognize $R = Q^T A$:

$$\begin{aligned} R = Q^T A &= \begin{bmatrix} \underline{\mathbf{q}}_1^T \\ \underline{\mathbf{q}}_2^T \\ \underline{\mathbf{q}}_3^T \end{bmatrix} \begin{bmatrix} \underline{\mathrm{a}}_1 & \underline{\mathrm{a}}_2 & \underline{\mathrm{a}}_3 \end{bmatrix} \\ &= \begin{bmatrix} \underline{\mathbf{q}}_1^T \underline{\mathbf{a}}_1 & \underline{\mathbf{q}}_1^T \underline{\mathbf{a}}_2 & \underline{\mathbf{q}}_1^T \underline{\mathbf{a}}_3 \\ \underline{\mathbf{q}}_2^T \underline{\mathbf{a}}_1 & \underline{\mathbf{q}}_2^T \underline{\mathbf{a}}_2 & \underline{\mathbf{q}}_2^T \underline{\mathbf{a}}_3 \\ \underline{\mathbf{q}}_3^T \underline{\mathbf{a}}_1 & \underline{\mathbf{q}}_3^T \underline{\mathbf{a}}_2 & \underline{\mathbf{q}}_3^T \underline{\mathbf{a}}_3 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \underline{\mathbf{q}}_1^T \underline{\mathbf{a}}_2 & \underline{\mathbf{q}}_1^T \underline{\mathbf{a}}_3 \\ 0 & \underline{\mathbf{q}}_2^T \underline{\mathbf{a}}_2 & \underline{\mathbf{q}}_2^T \underline{\mathbf{a}}_3 \\ 0 & 0 & \underline{\mathbf{q}}_3^T \underline{\mathbf{a}}_3 \end{bmatrix} \end{aligned}$$

Example 1: Given the 3×2 "tall" matrix:

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 12 \\ 6 & 10 \end{pmatrix}$$

obtain the QR-factorization.

- Consider the size of R: A is 3×2, so Q will be 3×2 and R will be 2×2.
- Implement least-squares method:
  - Set $\mathbf{q}_1$ as $\mathbf{x}_1$:

$$\mathbf{q}_1 = \begin{pmatrix} 2 \\ 3 \\ 6 \end{pmatrix}$$

  - Find $\mathbf{q}_2$:

$$\underline{\mathbf{c}}_2 - \frac{\underline{\mathbf{q}}_1^T \underline{\mathbf{c}}_2}{\underline{\mathbf{q}}_1^T \underline{\mathbf{q}}_1} \underline{\mathbf{q}}_1 = \begin{pmatrix} 1 \\ 12 \\ 10 \end{pmatrix} - \frac{2+36+60}{4+9+36} \begin{pmatrix} 2 \\ 3 \\ 6 \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \\ -2 \end{pmatrix}$$

  - Normalization:

$$Q = \frac{1}{7} \begin{pmatrix} 2 & -3 \\ 3 & 6 \\ 6 & -2 \end{pmatrix}$$

- Find R by using $R = Q^T A$:

$$Q^T A = Q^T(QR) = R$$

$$\Longrightarrow \begin{pmatrix} \underline{\mathbf{q}}_1^T \\ \underline{\mathbf{q}}_2^T \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 \mathbf{c}_2 \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{q}}_1^T \mathbf{c}1 & \underline{\mathbf{q}}_1^T \underline{\mathbf{c}}_2 \\ 0 & \underline{\mathbf{q}}_2^T \underline{\mathbf{c}}_2 \end{pmatrix} = \begin{pmatrix} 7 & 14 \\ 0 & 7 \end{pmatrix}$$

The critical part of this chapter. Recall how we the method of least-squares relied on a matrix A with independent columns:

$$A^T A\underline{\hat{x}} = A^T\underline{\mathbf{b}}$$

Now, with $A = QR$ - factorization we can write:

$$A^T A = (QR)^T QR = \left(R^T Q^T\right)(QR) = R^T \left(Q^T Q\right) R = R^T R$$

as $Q^T Q = I$, thus the problem becomes the need to solve:

$$R^T R\underline{\hat{\mathbf{x}}} = R^T Q^T \underline{\mathbf{b}}$$

Having established earlier that $A^T A$ is invertible, and therefore $R^T R = A^T A$ is also invertible. This means that we can multiply by $(R^T)^{-1}$:

$$\begin{aligned} R\underline{\hat{x}} &= Q^T\underline{\mathbf{b}} \\ \widehat{\boldsymbol{x}} &= R^{-1} Q^{\mathrm{T}} \boldsymbol{b} \end{aligned}$$

5 Eigenvalues, Eigenvectors and Diagonalisation

The first part was about $Ax = b$: balance and equilibrium and steady state. Now the second part is about change i.e. time enters the picture-continuous time in a differential equation or time steps in a difference equation. Those equations cannot be solved by elimination.

The key idea to solving these equations is to avoid all the complications presented by the matrix A. We want "eigenvectors" $\mathbf{x}$ that don't change direction when you multiply by A. A good example comes from the powers A, A^2, A^3,... of a matrix.

Suppose you need A^{100}. Its columns are very close to the eigenvector $(.6, .4)$:

$$A, A^2, A^3 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \quad \begin{bmatrix} .70 & .45 \\ .30 & .55 \end{bmatrix} \quad \begin{bmatrix} .650 & .525 \\ .350 & .475 \end{bmatrix} \qquad \boldsymbol{A^{100}} \approx \begin{bmatrix} \mathbf{.6000} & \mathbf{.6000} \\ \mathbf{.4000} & \mathbf{.4000} \end{bmatrix}$$

A^{100} was found by using the *eigenvalues* of A, not by multiplying 100 matrices. Those eigenvalues (here they are $\lambda = 1$ and $1/2$) are a new way to see into the heart of a matrix.

5.1 Eigenvectors

Almost all vectors change direction when multiplied by A. Certain exceptional vectors $\mathbf{x}$ are in the same direction as $A\mathbf{x}$. Those are the "eigenvectors" - Multiply an eigenvector by A, and the vector $A\mathbf{x}$ is a number λ times the original $\mathbf{x}$.

The basic equation is $A\mathbf{x} = \lambda\mathbf{x}$ where λ is an eigenvalue of A

The eigenvalue λ tells whether the special vector $\mathbf{x}$ is stretched or shrunk or reversed or left unchanged when multiplied by A. Recall that a λ value of 0 then $A\mathbf{x} = O\mathbf{x}$ means that this eigenvector is in the nullspace.

5.2 Solving the Eigenvalue Equation

Characteristic polynomial of the matrix A:

$$P_A(\lambda) = \det(A - \lambda I)$$

Characteristic equation:

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

This gives the algorithm: to find the eigenvalues and eigenvectors of a matrix A:

1. Fnd all scalars λ that are zeros of the characteristic equation to find the eigenvalues:

$$P_A(\lambda) = \det(A - \lambda I) = 0$$

2. For each such λ, find a non-zero solution of the equation to derive the corresponding eigenvectors:

$$(A - \lambda I)\underline{\mathbf{x}} = \underline{\mathbf{0}}$$

Example 1: Find eigenvalues and eigenvectors of the matrix

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

- Begin with the characteristic equation:

$$\begin{aligned} A - \lambda I = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} \\ \Longrightarrow \det(A - \lambda I) &= (2-\lambda)(2-\lambda) - 1 = 0 \end{aligned}$$

  solutions $\lambda = 3$ and 1.

- To find the eigenvectors, solve $(A - \lambda I)\underline{\mathbf{x}} = \underline{\mathbf{0}}$ for each eigenvalue.

  - $\lambda = 3$:

$$A - 3I = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\underline{\mathbf{x}} = \underline{\mathbf{0}}$$

    thus

$$\underline{\mathbf{x}}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

  - $\lambda = 1$:

$$A - I = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\underline{\mathbf{x}} = \underline{\mathbf{0}}$$

    thus

$$\underline{\mathbf{x}}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Example 2: Consider a rotation by $\frac{\pi}{2}$, with matrix

$$R = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

and find its eigenvalues and eigenvectors.

- $P_A(\lambda) = \det(A - \lambda I) = 0$:

$$\begin{aligned} P_A(\lambda) &= \det(A - \lambda I) \\ &= \begin{pmatrix} -\lambda & -1 \\ -1 & \lambda \end{pmatrix} \\ &= \lambda^2 + 1 \\ &\Rightarrow \lambda = \pm i \end{aligned}$$

- $\lambda = i$:

$$\begin{aligned} R - iI &= \left(\begin{array}{cc} -i & -1 \\ 1 & -i \end{array} \right) \Rightarrow \left(\begin{array}{cc} -i & -1 \\ 1 & -i \end{array} \right) \underline{\mathbf{x}} = \underline{\mathbf{0}} \\ \underline{\mathbf{x}}_1 &= \left(\begin{array}{c} i \\ 1 \end{array} \right) \end{aligned}$$

- $\lambda = -i$:

$$\begin{aligned} (R + iI)\underline{\mathbf{x}} &= \left(\begin{array}{cc} i & -1 \\ 1 & i \end{array} \right) \underline{\mathbf{x}} = \underline{\mathbf{0}} \\ \underline{\mathbf{x}}_2 &= \left(\begin{array}{c} 1 \\ i \end{array} \right) \end{aligned}$$

5.3 Diagonlisation

When x is an eigenvector, multiplication by A is just multiplication by a number λ i.e. $A\mathbf{x} = \lambda\mathbf{x}$. All the difficulties of matrices are swept away. Instead of an interconnected system, we can follow the eigenvectors separately. The 100th power of a diagonal matrix is easy.

The matrix A turns into a diagonal matrix $\mathbf{\Lambda}$ when eigenvectors are used properly. This is the matrix form of the previous chapter.

Diagonalization: Suppose the $[n \times n]$ matrix A has n linearly independent eigenvectors $x_1, ..., x_n$. Put them into the columns of an eigenvector matrix X. Then $X^{-1}AX$ is the eigenvalue matrix $\mathbf{\Lambda}$:

$$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \mathbf{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda_1} & & \\ & \ddots & \\ & & \boldsymbol{\lambda_n} \end{bmatrix}$$

This is the diagonalization of A and is available for most matrices.

5.4 Powers of matrices and the exponential matrix

The simplest use of diagonalization is in obtaining powers of a matrix in a manageable form.

Consider $A = S\mathbf{\Lambda}S^{-1}$ then:

$$\begin{aligned} A^2 &= (S\mathbf{\Lambda}S^{-1})(S\mathbf{\Lambda}S^{-1}) \\ &= S\mathbf{\Lambda}(S^{-1}S)\mathbf{\Lambda}S^{-1} \\ &= S\mathbf{\Lambda}^2S^{-1} \end{aligned}$$

Thus this can be summarised as:

$$A^k = S\mathbf{\Lambda}^kS^{-1}$$

Note that the same concept can be applied to exponentials:

$$e^A = Se^{\mathbf{\Lambda}}S^{-1}$$

As each of the exponential series on the diagonal in $e^{\mathbf{\Lambda}}$ converges, this guarantees convergence of the infinite series of matrices in e^A. This result is important in that it has many applications for example in systems of coupled ODEs.

5.5 Coupled First-Order ODEs

The matrix exponential can be applied to systems of coupled first-order equations. In systems control, we often come across such equations:

$$\dot{x}(t) = Ax(t) + Bu(t) \;\text{ where } x(t_0) = x_0$$

The solution commonly takes the form:

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau$$

Note that the equation actually represents a **state space model** which describes the behaviour in a particular space. Thus, $x(t)$ could be a vector of state variables $x = (x_1, x_2, \ldots, x_n)$ and A becoming a $[n \times n]$ matrix.

For example, consider two state variables $x_1(t)$ and $x_2(t)$. These two state variables has the following corresponding equations;

$$\begin{aligned}\dot{x}_1 &= ax_1 + bx_1 + u_1(t)\\ \dot{x}_2 &= ax_2 + bx_2 + u_2(t)\end{aligned}$$

By setting $u = 0$ to further simplify the equation:

$$\dot{\mathbf{x}} = \begin{pmatrix}\dot{x}_1\\ \dot{x}_2\end{pmatrix} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = A\mathbf{x}$$

Assuming A can be diagonalized i.e. two, independent eigenvectors, we will have an invertible matrix S and a diagonal matrix $\mathbf{\Lambda}$ such that:

$$A = S\mathbf{\Lambda}S^{-1} \;\text{ and }\; \mathbf{\Lambda} = S^{-1}AS$$

We now introduce a new vector $\mathbf{y(t)}$ which contains the solutions to the state variables e.g. $x_1(t)$. These new vector variables will satisfy the equation $\dot{y} = \mathbf{\Lambda y}$.

These vectors are related. Consider the substitution $\mathbf{x} = S\mathbf{y}$:

$$\begin{aligned}\dot{x} &= S\dot{y}\\ &= S(\mathbf{\Lambda y})\\ &= S(S^{-1}AS)\mathbf{y}\\ &= A(S\mathbf{y})\\ &= A\mathbf{x}\end{aligned}$$

This means that $\mathbf{x} = S\mathbf{y}$ is the solution to the system which only requires finding the eigenvalues and eigenvectors of A.

Example 1: Solve the system $\dot{\mathbf{x}} = A\mathbf{x}$ where $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ and $x_1(0) = x_2(0) = 1$.

- Note that eigenvalues and eigenvectors are assumed to have been found already.

$$\lambda_1 = 3 \text{ and } \lambda_2 = -1$$

  and

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \mathbf{a}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

- From the eigenvalues and eigenvectors, S and $\mathbf{\Lambda}$ is found:

$$S = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \text{ and } \mathbf{\Lambda} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$$

- Solve $\dot{\mathbf{y}} = \mathbf{\Lambda}\mathbf{y}$:

$$\underline{\dot{\mathbf{y}}} = \Lambda\underline{\mathbf{y}} \Rightarrow \begin{matrix} \dot{y_1} = 3y_1 \\ \dot{y_2} = -y_2 \end{matrix} \Rightarrow \underline{\mathbf{y}} = \begin{matrix} y_1 = c_1e^{3t} \\ y_2 = c_2e^{-t} \end{matrix}$$

- Expressed as a solution:

$$\underline{\mathbf{x}} = S\underline{\mathbf{y}} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} c_1e^{3t} \\ c_2e^{-t} \end{pmatrix} = \begin{pmatrix} c_1e^{3t} + c_2e^{-t} \\ 2c_1e^{3t} - 2c_2e^{-t} \end{pmatrix}$$

6 Symmetric Matrices and the SVD

Symmetric matrices S are the most important matrices the world will ever see in the theory of linear algebra and also in the applications. This is because:

- A symmetric matrix has only real eigenvalues.
- It is possible to find a full set of orthogonal eigenvectors

Spectral theorem: Every symmetric matrix has the factorization $S = QAQ^T$ with real eigenvalues in A and orthonormal eigenvectors in the columns of Q:

$$S = Q\boldsymbol{\Lambda}Q^{-1} = Q\boldsymbol{\Lambda}Q^T \quad \text{where } Q^{-1} = Q^T$$

Example 1: Fnd the orthogonal diagonalization of $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

- This matrix has the eigenvalues and corresponding eigenvectors of:

$$\lambda_1 = 1, \underline{\mathbf{x}}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \lambda_2 = 3, \underline{\mathbf{x}}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

  Note that A is symmetric, and the eigenvectors corresponding to distinct eigenvalues are automatically orthogonal.

- Normalize:

$$\begin{aligned} \hat{x}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} &\quad \text{and} \quad \hat{x}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \implies S = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} &\quad \text{and} \quad \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \end{aligned}$$

- The orthogonal diagonalization is:

$$\begin{aligned} A &= S\Lambda S^T \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \end{aligned}$$

6.1 The DFT Matrix

Hermitian: A matrix A is called Hermitian if

$$A^H = \bar{A}^T = A$$

Note that $\bar{A}^T$ is called the **hermitian transpose**.

The hermitian transpose can be found in the following:

- Given a matrix A:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2-i & 5 \\ 1+i & i & 4-2i \end{bmatrix}$$

- Transpose the matrix A:

$$\boldsymbol{A}^\top = \begin{bmatrix} 1 & 1+i \\ -2-i & i \\ 5 & 4-2i \end{bmatrix}$$

- Conjugate every entry of A:

$$\boldsymbol{A}^\mathrm{H} = \begin{bmatrix} 1 & 1-i \\ -2+i & -i \\ 5 & 4+2i \end{bmatrix}$$

For example, the matrix

$$A = \begin{pmatrix} 1 & 1-i \\ 1+i & 2 \end{pmatrix}$$

is Hermitian since $\bar{A}^T$ and A are the same.

Building on, if we have an **orthonormal** set, we can form the matrix Q which can be multiplied with the transpose of the complex conjugate.

$$Q^H Q \ = \ \bar{Q}^T Q = \begin{pmatrix} \bar{\underline{\mathbf{q}}}_1^T \\ \bar{\underline{\mathbf{q}}}_2^T \\ \vdots \\ \bar{\underline{\mathbf{q}}}_n^T \end{pmatrix} \begin{pmatrix} \underline{\mathbf{q}}_1 \, \underline{\mathbf{q}}_2 \cdots \underline{\mathbf{q}}_n \end{pmatrix} \begin{pmatrix} \bar{\underline{\mathbf{q}}}_1^T \underline{\mathbf{q}}_1 & \bar{\underline{\mathbf{q}}}_1^T \underline{\mathbf{q}}_2 & \cdots & \bar{\underline{\mathbf{q}}}_1^T \underline{\mathbf{q}}_n \\ \bar{\underline{\mathbf{q}}}_2^T \underline{\mathbf{q}}_1 & \bar{\underline{\mathbf{q}}}_2^T \underline{\mathbf{q}}_2 & \cdots & \bar{\underline{\mathbf{q}}}_2^T \underline{\mathbf{q}}_n \\ \vdots & \vdots & & \vdots \\ \bar{\underline{\mathbf{q}}}_n^T \underline{\mathbf{q}}_1 & \bar{\underline{\mathbf{q}}}_n^T \underline{\mathbf{q}}_2 & \cdots & \bar{\underline{\mathbf{q}}}_n^T \underline{\mathbf{q}}_n \end{pmatrix} \ = \ I_n$$

Unitary: An invertible complex matrix satisfying $\bar{Q}^T Q = I_n$

An extremely useful unitary matrix is the DFT matrix, also called Fourier matrix.

Let ω be a complex number satisfying $\omega^n = 1$, so that $1, \omega, \omega^2, \omega^3, \ldots \omega^{n-1}$ are the n^{th} roots of unity, and define

$$F_n = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \ldots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \ldots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \ldots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \ldots & \omega^{(n-1)(n-1)} \end{pmatrix}$$

For example, if $n = 4$, then $\omega = i$, $\omega^2 = -1$, $\omega^3 = -i$, and $\omega^4 = 1$, giving

$$F_4 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

and it's not hard to see that the columns are orthonormal vectors and that F_4 is in fact unitary.

6.2 Positive-definite Matrices

Positive-definite: A symmetric matrix whose eigenvalues are all positive.

A positive definite matrix needs to satisfy:

$$\underline{\mathbf{x}}^T \left(A^T A\right) \underline{\mathbf{x}} > 0 \quad \text{for every} \quad \underline{\mathbf{x}} \neq \underline{\mathbf{0}}$$

6.3 The Singular Value Decomposition (SVD)

The SVD is available for any matrix, including non-square unlike LU and QR.

Given any $m \times n$ matrix A, the Singular Value Decomposition of A is

$$A = U \Sigma V^T$$

where U and V are orthogonal matrices (unitary, if complex numbers involved) and Σ is diagonal, though not necessarily square. Σ will be $m \times n$ and will have zero-entries everywhere except on the main diagonal, where $i = j$.

Note the sizes of each matrix:

$$\underbrace{A}_{m \times n} = \underbrace{U}_{m \times m} \; \underbrace{\Sigma}_{m \times n} \; \underbrace{V^T}_{n \times n}$$

Example 1: Find the SVD of $A = \begin{pmatrix} 2 & 2 & 0 \\ -1 & 1 & 0 \end{pmatrix}$

- Note the rank: $r = 2$ - two **singular values** required
- Note the matrix size: 2×3 matrix thus:
  - $U = 2 \times 2$
  - $\Sigma = 2 \times 3$
  - $V = 3 \times 3$
- Find AA^T:

$$A^T A = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$$

- Find the eigen values and vectors:
  - Eigenvalues: 8 and 2
  - Eigenvectors:

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- List the singular values: $\sqrt{8}$ and $\sqrt{2}$
- Find the V vector:

6.4 Pseudoinverse

If A^{-1} does not exist for a given A, a pseudoinverse A^{+} can be obtained by using the SVD.

$$\underbrace{A^{+}}_{n\times m} = \underbrace{V}_{n\times n}\underbrace{\Sigma^{+}}_{n\times m}\underbrace{U^{T}}_{m\times m} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_r} & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \left(\underline{\mathbf{u}}_1\underline{\mathbf{u}}_2 \dots \underline{\mathbf{u}}_r\underline{\mathbf{u}}_{r+1} \dots \underline{\mathbf{u}}_m\right)^{T}$$

When A is invertible, then A^{+} is the same as the inverse, but when not, we can still multiply

$$AA^{+} = \left(U\Sigma V^{T}\right)\left(V\Sigma^{+}U^{T}\right) = U\left(\Sigma\Sigma^{+}\right)U^{T}$$

where $\Sigma\Sigma^{+}$ is an $m \times m$ diagonal matrix with r ones on the diagonal and $m-r$ zeroes. Multiplying by U on the left and U^{T} on the right leaves these in place and we end up with

$$AA^{+} = \Sigma\Sigma^{+} = \begin{pmatrix} I_r & 0 \\ 0 & \underline{0} \end{pmatrix}$$

Note AA^{+} and $A^{+}A$ are different:

$$A^{+}A = \left(V\Sigma^{+}U^{T}\right)\left(U\Sigma V^{T}\right) = V\Sigma^{+}\Sigma V^{T} = \Sigma^{+}\Sigma = \begin{pmatrix} I_r & 0 \\ 0 & \underline{0} \end{pmatrix}$$

The amount of zeroes on the diagonal in each case depends on r, m, n.

Example 4.3. *Find the pseudoinverse of* $A = \begin{pmatrix} 2 & 2 & 0 \\ -1 & 1 & 0 \end{pmatrix}$.

We have

$$A^{+} = V\Sigma^{+}U^{T} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{8}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}$$

and multiplying

$$AA^{+} = \begin{pmatrix} 2 & 2 & 0 \\ -1 & 1 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{4} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

because $r = m = 2$, whereas

$$A^{+}A = \begin{pmatrix} \frac{1}{4} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & 2 & 0 \\ -1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

because $r = 2 < 3 = n$.