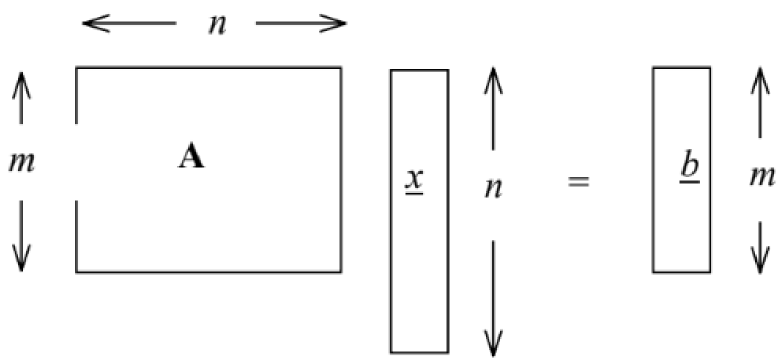
The Geometry of n Dimensions

In general, an $m \times n$ matrix $\mathbf{A}$, transforms an n-dimensional vector $\underline{x}$ into a corresponding m-dimensional vector $\underline{b}$:

$\mathbf{A}\underline{x} = \underline{b}$



Vector Spaces and Subspaces

For a general non-square $m \times n$ matrix:

$$\mathbf{A}\underline{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\underline{a}_1 + x_2\underline{a}_2 + \cdots + x_n\underline{a}_n$$

The $\underline{a}_i$ are the **column vectors** of $\mathbf{A}$ and they sweep out the part of R^m that we can get to by multiplying a vector in R^n by $\mathbf{A}$.

The region mapped out in R^m as the x_i vary is called a **vector space** and it is a straightforward generalisation to arbitrary dimensions of the concept of a line or a plane.

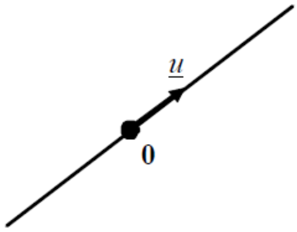
A vector space in R^m is the set of $\underline{x}$ of the form:

$\underline{x} = \lambda\underline{u} + \mu\underline{v} + \cdots$

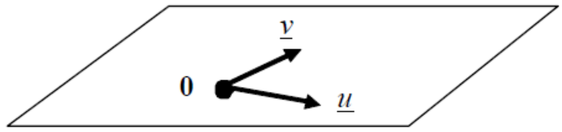
R^m is itself a vector space and "smaller" ones within it are said to be sub-spaces of R^m.

For example: The non-trivial sub-spaces of R^3 are

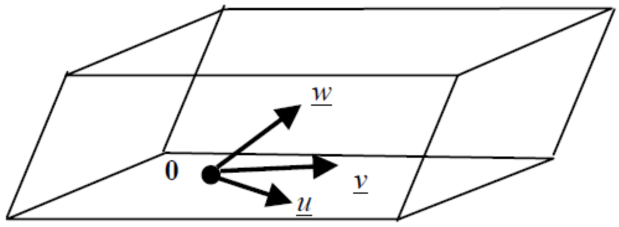
(a) $\underline{x} = \lambda \underline{u}$ lines through the origin



(b) $\underline{x} = \lambda \underline{u} + \mu \underline{v}$ planes through the origin.



(c) $\underline{x} = \lambda \underline{u} + \mu \underline{v} + \nu \underline{w}$ = whole of R^3



Column Space

The vector space spanned by the columns of a general $m \times n$ matrix $\mathbf{A}$ is called the **column space** of
$\mathbf{A}$. The dimension (in the degrees of freedom sense) of column space is called the **rank** of $\mathbf{A}$.

> Column space is part of R^m and so the number of independent columns of $\mathbf{A}$ can not exceed m, i.e. $\text{rank}(A) \leq m$. In addition, there are only n columns, so $\text{rank}(A) \leq n$.

> If $\underline{b}$ lies in column space, then $\mathbf{A}\underline{x} = \underline{b}$ has at least one solution.

> If $\underline{b}$ is not in column space, then $\mathbf{A}\underline{x} = \underline{b}$ has no solution.

> If $\text{rank}(A) = m$, so that column space = whole of R^m, then $\underline{b}$ must lie in column space.

Matrix Multiplication

For the general case of $\mathbf{A} = \mathbf{BC}$ where $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{B}$ is an $m \times k$ matrix, $\mathbf{C}$ is an $k \times n$ matrix,

$$\begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \underline{b}_1 & \underline{b}_2 & \cdots & \underline{b}_k \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kn} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \uparrow \\ \underline{b}_1 \\ \downarrow \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \end{bmatrix} + \begin{bmatrix} \uparrow \\ \underline{b}_2 \\ \downarrow \end{bmatrix} \begin{bmatrix} c_{21} & c_{22} & \cdots & c_{2n} \end{bmatrix} + \cdots + \begin{bmatrix} \uparrow \\ \underline{b}_k \\ \downarrow \end{bmatrix} \begin{bmatrix} c_{k1} & c_{k2} & \cdots & c_{kn} \end{bmatrix}$$

If we denote the columns of $\mathbf{B}$ as $\underline{b}_i$ and the rows of $\mathbf{C}$ as $\underline{\tilde{c}}_i$:

$\mathbf{A} = \underline{b}_1 \underline{\tilde{c}}_1^T + \underline{b}_2 \underline{\tilde{c}}_2^T + \cdots + \underline{b}_k \underline{\tilde{c}}_k^T$

A product of vectors $\underline{x}\underline{y}^T$ is referred to as an **outer product** of $\underline{x}$ and $\underline{y}$. (The dot product $\underline{x}^T \underline{y} = \underline{x} \cdot \underline{y}$ is also called an **inner product**). So that $\mathbf{A}$ is the sum of the outer products of each of the columns of $\mathbf{B}$ with the corresponding row of $\mathbf{C}$.

For the j-th column of $\mathbf{A}$:

$$\begin{bmatrix} \uparrow \\ \underline{a}_j \\ \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow \\ \underline{b}_1 \\ \downarrow \end{bmatrix} [\, c_{1j} \,] + \begin{bmatrix} \uparrow \\ \underline{b}_2 \\ \downarrow \end{bmatrix} [\, c_{2j} \,] + \cdots + \begin{bmatrix} \uparrow \\ \underline{b}_k \\ \downarrow \end{bmatrix} [\, c_{kj} \,]$$

The matrix multiplication can also be interpreted as:

$$\begin{bmatrix} \leftarrow & \underline{\tilde{a}}_1 & \rightarrow \\ \leftarrow & \underline{\tilde{a}}_2 & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \underline{\tilde{a}}_m & \rightarrow \end{bmatrix} = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{m1} \end{bmatrix} [\leftarrow \quad \underline{\tilde{c}}_1 \quad \rightarrow] + \begin{bmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{m2} \end{bmatrix} [\leftarrow \quad \underline{\tilde{c}}_2 \quad \rightarrow] + \cdots + \begin{bmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{mk} \end{bmatrix} [\leftarrow \quad \underline{\tilde{c}}_k \quad \rightarrow]$$

For the i-th row of $\mathbf{A}$:

$$[\leftarrow \quad \underline{\tilde{a}}_i \quad \rightarrow] = [\, b_{i1} \,] [\leftarrow \quad \underline{\tilde{c}}_1 \quad \rightarrow] + [\, b_{i2} \,] [\leftarrow \quad \underline{\tilde{c}}_2 \quad \rightarrow] + \cdots + [\, b_{ik} \,] [\leftarrow \quad \underline{\tilde{c}}_k \quad \rightarrow]$$

LU Factorisation

An $m \times n$ matrix $\mathbf{A}$ can be factorized into the form $\mathbf{A} = \mathbf{LU}$ where $\mathbf{L}$ is **lower-triangular** $m \times m$ matrix with 1's down the leading diagonal and $\mathbf{U}$ is an **upper-echelon** matrix which is the same shape as $\mathbf{A}$.

Lower triangular means that $\mathbf{L}$ has non-zero terms only on and below the leading diagonal:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 1 \end{bmatrix}$$

Upper echelon means that all non-zero elements are on or above the leading diagonal:

$$\mathbf{U} = \begin{bmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{bmatrix}$$

If $\mathbf{A}$ is square, then so is $\mathbf{U}$ which is then said to be **upper triangular**.

$$\mathbf{A} = \underline{l}_1 \underline{\tilde{u}}_1^T + \underline{l}_2 \underline{\tilde{u}}_2^T + \cdots + \underline{l}_k \underline{\tilde{u}}_k^T$$

Solution to Matrix Equation

$$\mathbf{A}\underline{x} = (\mathbf{LU}\underline{x}) = \mathbf{L}(\mathbf{U}\underline{x}) = \underline{b}$$

$\mathbf{A}\underline{x} = \underline{b}$ is solved in two steps, find $\underline{c}$ from $\mathbf{L}\underline{c} = \underline{b}$ and then find $\underline{x}$ from $\mathbf{U}\underline{x} = \underline{c}$.

Partial pivoting

The technique of scanning the remaining non-zero elements in the next column in the remainder matrix to be zeroed and choosing the largest (in absolute value) at every stage is called **partial pivoting**.

LU decomposition with partial pivoting is $\mathbf{PA} = \mathbf{LU}$ where $\mathbf{P}$ is the permutation matrix. LU in its partial-pivoting mode is immune to introducing ill-conditioning into a problem.

Bases for the Column Space and Row Space

Column Space = all vectors formed by taking a linear combination of the columns of $\mathbf{A}$:

$$\lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \lambda_n \underline{a}_n$$

Row Space = all vectors formed by taking a linear combination of the rows of $\mathbf{A}$:

$$\mu_1 \underline{\tilde{a}}_1 + \mu_2 \underline{\tilde{a}}_2 + \cdots + \mu_m \underline{\tilde{a}}_m$$

$\underline{l}_i$ form a basis for the column space of $\mathbf{A}$ (the set of $\underline{l}_i$ for which the corresponding $\underline{\tilde{u}}$ is non-zero).

$$\begin{bmatrix} \uparrow \\ \underline{a}_j \\ \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow \\ \underline{l}_1 \\ \downarrow \end{bmatrix} [u_{1j}] + \begin{bmatrix} \uparrow \\ \underline{l}_2 \\ \downarrow \end{bmatrix} [u_{2j}] + \cdots + \begin{bmatrix} \uparrow \\ \underline{l}_m \\ \downarrow \end{bmatrix} [u_{mj}]$$

$$\underline{a}_j = u_{1j} \underline{l}_1 + u_{2j} \underline{l}_2 + \cdots + u_{mj} \underline{l}_m$$

$\underline{\tilde{u}}_i$ form a basis for the row space of $\mathbf{A}$ (the set of non-zero $\underline{\tilde{u}}_i$).

$$[\leftarrow \quad \underline{\tilde{a}}_i \quad \rightarrow] = [l_{i1}][\leftarrow \quad \underline{\tilde{u}}_1 \quad \rightarrow] + [l_{i2}][\leftarrow \quad \underline{\tilde{u}}_2 \quad \rightarrow] + \cdots + [l_{im}][\leftarrow \quad \underline{\tilde{u}}_m \quad \rightarrow]$$

$$\underline{\tilde{a}}_i = l_{i1} \underline{\tilde{u}}_1 + l_{i2} \underline{\tilde{u}}_2 + \cdots + l_{im} \underline{\tilde{u}}_m$$

> The dimension of column space is equal to the columns of $\mathbf{L}$ that correspond to non-zero rows of $\mathbf{U}$. The dimension of row space is also equal to the number of non-zero rows of $\mathbf{U}$. For any matrix, number of independent rows is equal to number of independent columns (rank of $\mathbf{A}$).

Properties of the L & U Matrices

The columns of $\mathbf{L}$ and the non-zero rows of $\mathbf{U}$ are independent.

$\mathbf{L}^{-1}$ always exists and $\det(\mathbf{L}) = 1$.

Since the column space of $\mathbf{L}$ is the whole of R^m, any vector in R^m can be expressed in terms of the columns of $\mathbf{L}$.

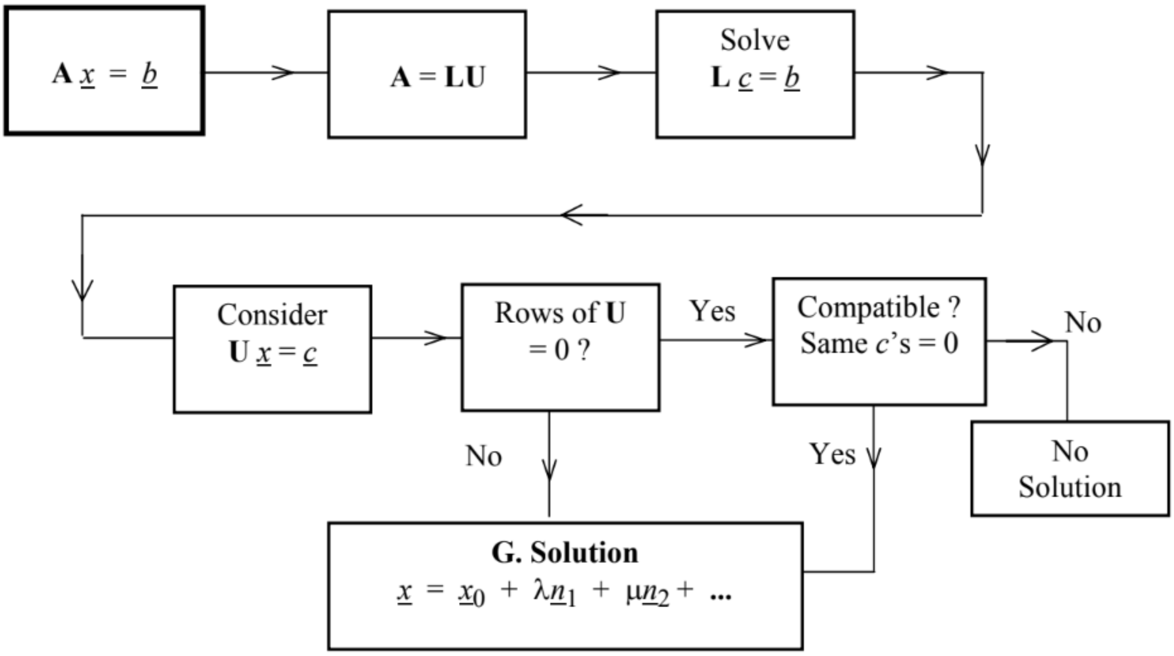
$\mathbf{U}$ is the same shape as $\mathbf{A}$ and $\mathbf{U}$ has an inverse if and only if $\mathbf{A}$ has one.

Algorithmic Complexity

For large n, the LU factorisation requires $\approx \frac{2}{3}n^3$ operations. Once the LU factorisation has been completed, it requires $\approx 2n^2$ operations to complete the solution of $\mathbf{A}\underline{x} = \underline{b}$.

The Solution of Matrix Equation

General Solution



> Set all free variables to zero and find a particular solution x_0.

> Set the RHS to zero, give each free variable in turn the value 1 while the others are zero, and solve to find a set of vectors which span the null space of $\mathbf{A}$.

Properties of the Fundamental Subspaces

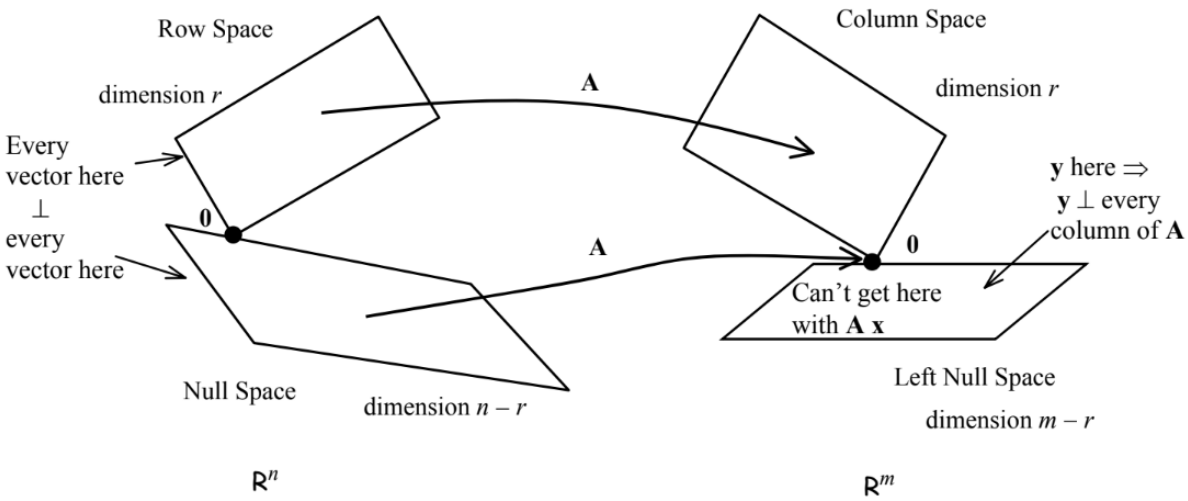
Bases for the Four Spaces

Column Space: the columns of $\mathbf{L}$ used, corresponding to non-zero rows of $\mathbf{U}$.

Null Space (same as the Null Space of $\mathbf{U}$): set the free variables to 1 in turn and solve $\mathbf{U}\underline{x} = 0$.

Row Space: all non-zero rows of $\mathbf{U}$.

Left Null Space: set the free variables to 1 in turn and solve $\mathbf{L}_{RED}^T\underline{b} = 0$.



Properties of the Fundamental Subspaces

> The dimension of row-space is equal to the number of non-zero rows of $\mathbf{U}$ which is equal to the number of basic variables (those with pivots).

$\text{Dimension of row space} = \text{dimension of column space} = r = \text{rank}(\mathbf{A})$

> The dimension of the null space of of $\mathbf{A}$ is equal to the number of free variables (the number of variables without pivots).

$\text{Dimension of null space} = n - r$

> A vector $\underline{n}$ is in null-space if, and only if, it is orthogonal to every row of $\mathbf{A}$ and hence orthogonal to every vector in row space. (Null space and Row Space are orthogonal)

$$\mathbf{A}\underline{n} = \begin{bmatrix} \leftarrow & \underline{\tilde{a}}_1 & \rightarrow \\ \leftarrow & \underline{\tilde{a}}_2 & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \underline{\tilde{a}}_m & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \underline{n} \\ \downarrow \end{bmatrix} = \begin{bmatrix} \underline{\tilde{a}}_1 \cdot \underline{n} \\ \underline{\tilde{a}}_2 \cdot \underline{n} \\ \vdots \\ \underline{\tilde{a}}_m \cdot \underline{n} \end{bmatrix} = 0$$

> Every vector in Column Space is orthogonal to every vector in Left Null Space.

$\text{Dimension of left null-space} = m - r$

Row space and Null Space are said to be **orthogonal complements** in R^n, and Column Space and Left Null space are orthogonal complements in R^m.

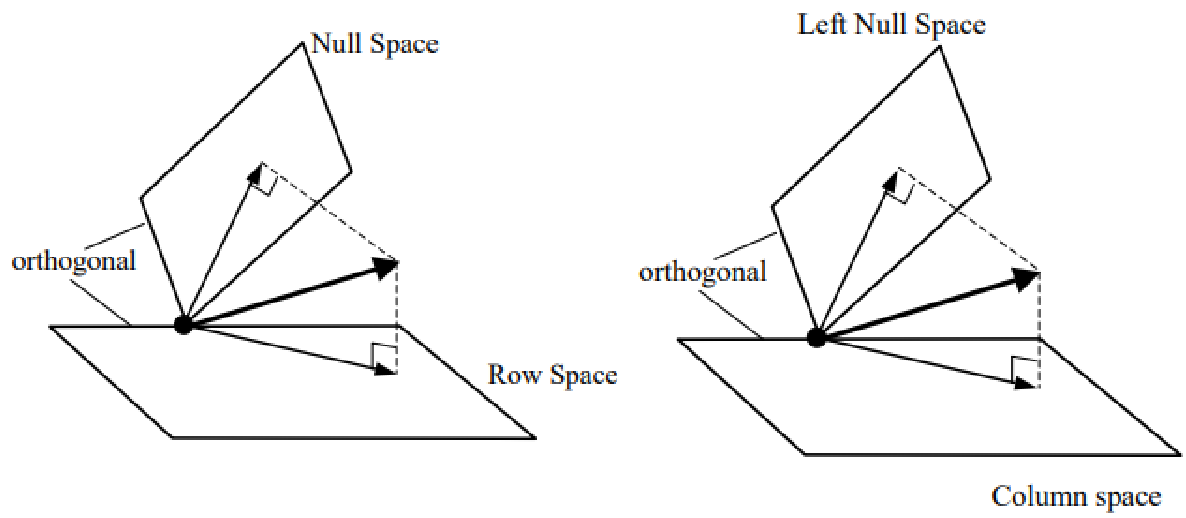
The Big Picture

Row space and Null space represent all of R^n. Column space and Left Null space similarly carve up R^m.

$\underline{x} = \underline{x}_{row} + \underline{x}_{null}$

$\underline{b} = \underline{b}_{col} + \underline{b}_{left}$

$\underline{x}_{row} \cdot \underline{x}_{null} = \underline{b}_{col} \cdot \underline{x}_{left} = 0$



If a vector $\underline{b}$ has a non-zero $\underline{b}_{left}$ then $\mathbf{A}\underline{x} = \underline{b}$ has no solution. The various compatibility conditions necessary for there to be a of solution can be stated as $\underline{b}$ being orthogonal to each of the vectors in a basis of Left Null Space.

If we restrict the vectors that $\mathbf{A}$ operates on to Row Space, then it is clear that $\mathbf{A}$ maps an r-dimensional space onto Column Space (which is also r-dimensional). The mapping is, therefore, reversible.

Least Squares Solution and QR Factorisation

Suppose we have carried out an experiment, in which the parameter b has been measured at different times t, and that we are seeking to fit a linear relationship to the data $b = C + Dt$ or for a quadratic fit $b = C + Dt + Et^2$.

In matrix form the linear case is:

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

These equations are **inconsistent** and we need to find **the best fit**: $\mathbf{A}\overline{\mathbf{x}}$ as close as possible to $\mathbf{b}$.

For least squares problems, then, the $m \times n$ matrix $\mathbf{A}$ usually has the following properties:

> $m > n$ (often $m \gg n$)

> The columns of $\mathbf{A}$ are independent. (rank of $\mathbf{A}$ is n.)

The least squares solution for $\mathbf{x}$ (= $\overline{\mathbf{x}}$) minimises $|\mathbf{Ax} - \mathbf{b}|^2 = (\mathbf{Ax} - \mathbf{b}) \cdot (\mathbf{Ax} - \mathbf{b})$ and this can be multiplied out and then partial differentiation used to find the minimum.

Alternatively, for any vector $\mathbf{b} = \mathbf{b}_{col} + \mathbf{b}_{left}$ where $\mathbf{b}_{col} \cdot \mathbf{b}_{left} = 0$. So we need to get rid of $\mathbf{b}_{left}$ and just concentrate on $\mathbf{b}_{col}$ by multiplying the original problem by $\mathbf{A}^t$.

$\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b} = \mathbf{A}^t\mathbf{b}_{col} + \mathbf{A}^t\mathbf{b}_{left} = \mathbf{A}^t\mathbf{b}_{col}$

In summary, the least squares solution to an inconsistent system $\mathbf{Ax} = \mathbf{b}$ of m equations in n unknowns satisfies $\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b}$.

Assuming that the columns of $\mathbf{A}$ are independent, $\mathbf{A}^t\mathbf{A}$ is invertible and $\overline{\mathbf{x}} = \left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{b}$

The Gram-Schmidt Process

Another way of removing the part of $\mathbf{b}$ that is not in the column space of $\mathbf{A}$ is to project $\mathbf{b}$ directly onto column space.

The **Gram-Schmidt procedure** is a way of generating a set of mutually orthogonal unit vectors (orthogonal + unit = orthonormal) from an arbitrary set. Armed with these, taking projections is much easier.

$\mathbf{b}_{col} = \alpha_1\mathbf{q}_1 + \alpha_2\mathbf{q}_2 + \cdots$

$\alpha_i = \mathbf{q}_i \cdot \mathbf{b}$

We start with the columns of $\mathbf{A}$ and derive the $\mathbf{q}$'s as follows:

1. Turn the first one into a unit vector

$\mathbf{q}_1 = \frac{\mathbf{a}_1}{|\mathbf{a}_1|}$

2. Take $\mathbf{a}_2$ and form $\mathbf{q}_2$ by first subtracting off the bit that's parallel to $\mathbf{a}_1$ and then normalising

$\tilde{\mathbf{a}}_2 = \mathbf{a}_2 - (\mathbf{q}_1 \cdot \mathbf{a}_2)\,\mathbf{q}_1$

$\mathbf{q}_2 = \frac{\tilde{\mathbf{a}}_2}{|\tilde{\mathbf{a}}_2|}$

3. Repeat this process for the other $\mathbf{a}$'s

$\tilde{\mathbf{a}}_i = \mathbf{a}_i - (\mathbf{q}_1 \cdot \mathbf{a}_i)\,\mathbf{q}_1 - (\mathbf{q}_2 \cdot \mathbf{a}_i)\,\mathbf{q}_2 - \cdots$

$\mathbf{q}_i = \frac{\tilde{\mathbf{a}}_i}{|\tilde{\mathbf{a}}_i|}$

QR Factorisation

If we assemble the vectors $\mathbf{a}_i$ from the previous section as the columns of a matrix $\mathbf{A}$, and vectors $\mathbf{q}_i$ as those of a matrix $\mathbf{Q}$, then we have $\mathbf{A} = \mathbf{QR}$.

$$\begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \cdot \mathbf{a}_1 & \mathbf{q}_1 \cdot \mathbf{a}_2 & \cdots & \mathbf{q}_1 \cdot \mathbf{a}_n \\ 0 & \mathbf{q}_2 \cdot \mathbf{a}_2 & \cdots & \mathbf{q}_2 \cdot \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{q}_n \cdot \mathbf{a}_n \end{bmatrix}$$

The columns of $\mathbf{Q}$ are mutually orthogonal vectors which span the column space of $\mathbf{Q}$.

$\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$

The matrix $\mathbf{R}$ is square, upper triangular with non-zero elements down the diagonal. It therefore has rank n and is invertible.

Simplification of Least Squares Solution

Given the set of equations $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is an $m \times n$ matrix whose columns are independent ($m \geq n$ and rank of $\mathbf{A}$ is n) then the least squares solution satisfies:

$\mathbf{A}^t\mathbf{A}\overline{\mathbf{x}} = \mathbf{A}^t\mathbf{b}$

$\mathbf{R}\overline{\mathbf{x}} = \mathbf{Q}^t\mathbf{b}$

Projection onto Column Space

The projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$:

$\mathbf{b}_{col} = \mathbf{A}\left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{b} = \mathbf{Pb}$

The projection matrix $\mathbf{P}$ is given by:

$\mathbf{P} = \mathbf{A}\left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t = \mathbf{QQ}^t$

$\mathbf{b}_{col} = (\mathbf{q}_1 \cdot \mathbf{b})\,\mathbf{q}_1 + (\mathbf{q}_2 \cdot \mathbf{b})\,\mathbf{q}_2 + \cdots + (\mathbf{q}_n \cdot \mathbf{b})\,\mathbf{q}_n$

$$\mathbf{b}_{col} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \cdot \mathbf{b} \\ \mathbf{q}_2 \cdot \mathbf{b} \\ \vdots \\ \mathbf{q}_n \cdot \mathbf{b} \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \underline{\mathbf{q}}_1 & \rightarrow \\ \leftarrow & \underline{\mathbf{q}}_2 & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \underline{\mathbf{q}}_n & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \mathbf{b} \\ \downarrow \end{bmatrix} = \mathbf{QQ}^t\mathbf{b}$$

Eigenvalues and Eigenvectors

For a symmetric matrix $\mathbf{A}$, the eigenvalues are real and the eigenvectors orthogonal.

$$\mathbf{A}\begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ u_1 & u_2 & \cdots & u_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\mathbf{A} = \mathbf{U\Lambda U}^{-1} = \mathbf{U\Lambda U}^{T}$$

$$\mathbf{A}^k = \left(\mathbf{U\Lambda U}^{-1}\right)\left(\mathbf{U\Lambda U}^{-1}\right)\cdots\left(\mathbf{U\Lambda U}^{-1}\right) = \mathbf{U\Lambda}^k\mathbf{U}^{-1}$$

Characteristic Equation

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

For there to be a nonzero solution for $\mathbf{x}$, $\mathbf{A} - \lambda\mathbf{I}$ is singular.

For a general $n \times n$ matrix $\mathbf{A}$, the **characteristic equation** takes the form:

$$\mathrm{P}(\lambda) = \det\left(\mathbf{A} - \lambda\mathbf{I}\right) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0$$

$$\mathrm{P}(\lambda) = (-\lambda)^n + \alpha_1(-\lambda)^{n-1} + \alpha_2(-\lambda)^{n-2} + \cdots + \alpha_n = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) = 0$$

The product of the eigenvalues is simply the determinant of $\mathbf{A}$.

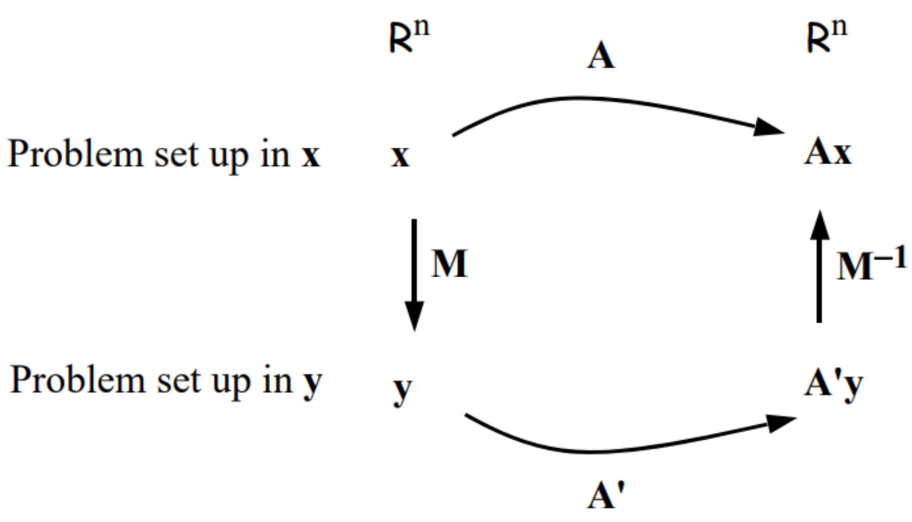
A matrix which does not have a full set of eigenvectors is said to be **defective**.

Diagonal Form of a Matrix

In general, the eigenvectors for non-symmetric matrices are not orthogonal. However, the columns of $\mathbf{X}$ are independent. This means that the column space of $\mathbf{X}$ has dimension n and therefore must be the whole of R^n. ($\mathbf{X}$ is invertible)

$$\mathbf{A} = \mathbf{X\Lambda X}^{-1}$$

A relationship of this form is referred to as a **similarity** relationship and $\mathbf{A}$ and $\mathbf{\Lambda}$ are said to be similar matrices.



$$\mathbf{Ax} = \mathbf{M}^{-1}\mathbf{A'y} \;=\; \mathbf{M}^{-1}\mathbf{A'Mx}$$

$\mathbf{A}$ is similar to $\mathbf{A}^{'}$ since $\mathbf{A} = \mathbf{M}^{-1}\mathbf{A}^{'}\mathbf{M}$ and $\mathbf{A}^{'} = \mathbf{MAM}^{-1}$

Similar matrices have the same set of eigenvalues and the eigenvectors are related by $\mathbf{y} = \mathbf{Mx}$.

A matrix which is similar to a diagonal matrix is said to be **diagonalisable**.

Diagonalisable (Non-Symmetric) Matrices

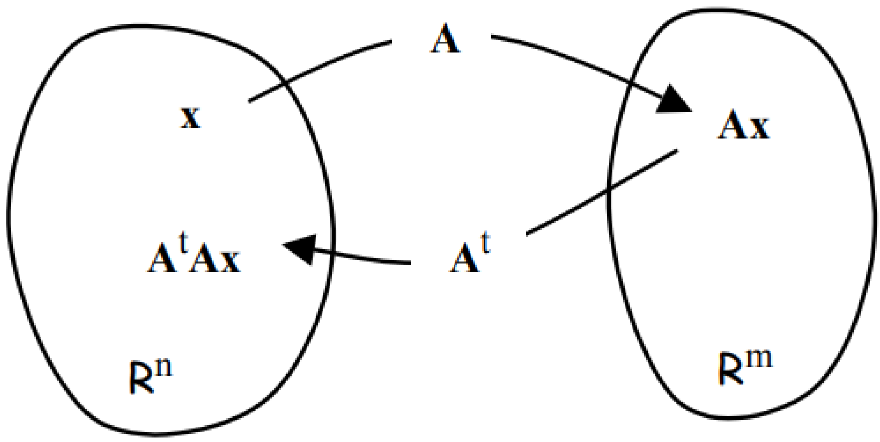
> Any matrix with distinct eigenvalues can be diagonalized.

Non-symmetric matrices with repeated eigenvalues may be diagonalisable, or they may not.

Singular Value Decomposition

Non-square matrices do not have eigenvalues and eigenvectors but there is a related concept called **singular values**.

The Eigenvalue Problem



If $\mathbf{A}$ is $m \times n$, then $\mathbf{A}^t$ is $n \times m$ and $\mathbf{A}^t\mathbf{A}$ is $n \times n$. It is also symmetric since $\left(\mathbf{A}^t\mathbf{A}\right)^t = \mathbf{A}^t\mathbf{A}$.

$\mathbf{A}^t\mathbf{A}$ is, therefore, diagonalisable and the eigenvectors of $\mathbf{A}^t\mathbf{A}$ can be chosen to be orthogonal, unit vectors.

$\mathbf{A}^t\mathbf{A} = \mathbf{Q\Lambda Q}^t$

The eigenvalues of $\mathbf{A}^t\mathbf{A}$ can not be negative.

$\mathbf{A}^t\mathbf{Aq} = \lambda\mathbf{q}$

$\lambda = \frac{(\mathbf{Aq})^t\mathbf{Aq}}{\mathbf{q}^t\mathbf{q}} = \frac{|\mathbf{Aq}|^2}{|\mathbf{q}|^2}$

> There will be r which are non-zero and $n - r$ which are zero where $r = \text{rank}(\mathbf{A})$.

For a zero eigenvalue, $\mathbf{Aq} = 0$ and so $\mathbf{q}$ is in the null space of $\mathbf{A}$.

For a non-zero eigenvalue, $\mathbf{Aq}$ is non-zero and so $\mathbf{q}$ is not in the null space of $\mathbf{A}$. But all the $\mathbf{q}$'s are orthogonal, so it must be orthogonal to all of the $\mathbf{q}$'s which are in null space, and hence must be in the row space of $\mathbf{A}$.

$$\mathbf{Q} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow & & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_r & \mathbf{q}_{r+1} & \mathbf{q}_{r+2} & \cdots & \mathbf{q}_n \\ \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\underbrace{\mathbf{q}_1 \cdots \mathbf{q}_r}_{\text{non-zero } \lambda} \quad \underbrace{\mathbf{q}_{r+1} \cdots \mathbf{q}_n}_{\lambda = 0}$$

$\mathbf{q}_1 \ldots \mathbf{q}_r$ are an orthogonal basis for Row Space

$\mathbf{q}_{r+1} \ldots \mathbf{q}_n$ are an orthogonal basis for Null Space

Singular Values

Define the **singular values** of $\mathbf{A}$ by $\sigma_i = \sqrt{\lambda_i}$ for $1 \leq i < r$. (non-zero eigenvalues of $\mathbf{A}^t\mathbf{A}$)

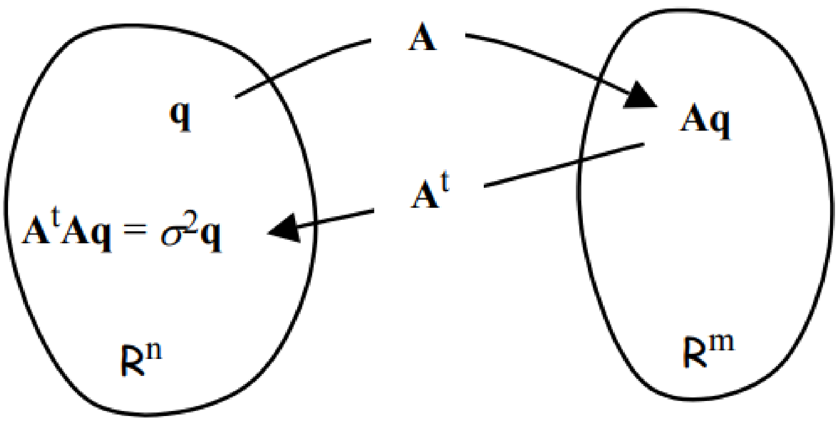
$\mathbf{A}^t\mathbf{A}\mathbf{q}_i = \sigma_i^2\mathbf{q}_i$

The **norm** of $\mathbf{A}$ (a measure of the size of the matrix, where a vector has a magnitude) is equal to the largest singular value of $\mathbf{A}$.

$\|\mathbf{A}\| = \sigma_{max}$

Basis for Column Space and Null Space

$\hat{\mathbf{q}}_i = \frac{\mathbf{A}\mathbf{q}_i}{\sigma_i}$



$\mathbf{q}$'s are orthogonal, unit vectors in the column space of $\mathbf{A}$. They are eigenvectors for $\mathbf{A}\mathbf{A}^t$ with the same eigenvalues σ_i^2.

$\mathbf{A}\left(\mathbf{A}^t\mathbf{A}\mathbf{q}\right) = \mathbf{A}\mathbf{A}^t\left(\mathbf{A}\mathbf{q}\right) = \sigma^2\mathbf{A}\mathbf{q}$

We can complete this set by adding eigenvectors corresponding to eigenvalue zero for $\mathbf{A}\mathbf{A}^t$. They must be in the left-null space of $\mathbf{A}$)and can be chosen to be orthogonal, unit vectors.

$\mathbf{A}\mathbf{A}^t = \hat{\mathbf{Q}}\hat{\Lambda}\hat{\mathbf{Q}}^t$

$$\hat{\mathbf{Q}} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow & & \uparrow \\ \hat{\mathbf{q}}_1 & \hat{\mathbf{q}}_2 & \cdots & \hat{\mathbf{q}}_r & \hat{\mathbf{q}}_{r+1} & \hat{\mathbf{q}}_{r+2} & \cdots & \hat{\mathbf{q}}_m \\ \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\underbrace{\hspace{6em}}_{\textbf{non-zero } \sigma} \qquad \underbrace{\hspace{6em}}_{\sigma = 0}$$

$\hat{\mathbf{q}}_1 \ldots \hat{\mathbf{q}}_r$ are an orthogonal basis for Column Space

$\hat{\mathbf{q}}_{r+1} \ldots \hat{\mathbf{q}}_m$ are an orthogonal basis for Left-Null Space

Singular Value Decomposition

Since $\mathbf{A}\mathbf{q}_i = \sigma_i \hat{\mathbf{q}}_i$ for $1 \leq i < r$, we can write $\mathbf{A}\mathbf{Q} = \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}$.

$$\mathbf{A} \begin{bmatrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ \mathbf{q}_1 & \cdots & \mathbf{q}_r & \mathbf{q}_{r+1} & & \mathbf{q}_n \\ \downarrow & & \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ \hat{\mathbf{q}}_1 & \cdots & \hat{\mathbf{q}}_r & \hat{\mathbf{q}}_{r+1} & \cdots & \hat{\mathbf{q}}_m \\ \downarrow & & \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & & \\ & \cdots & & & \mathbf{0} & & \\ & & \sigma_r & & & & \\ & & & 0 & & & \\ & \mathbf{0} & & & \cdots & & \\ & & & & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ \sigma_1\hat{\mathbf{q}}_1 & \cdots & \sigma_r\hat{\mathbf{q}}_r & \mathbf{0} & \cdots & \mathbf{0} \\ \downarrow & & \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Because $\mathbf{Q}$ is an orthogonal matrix, its inverse is its transpose.

$\mathbf{A} = \hat{\mathbf{Q}}\mathbf{\Sigma}\mathbf{Q}^t$