Imperial College London

Maths for Signals and Systems

Session 1 (Week 1-2)

Pier Luigi Dragotti
EEE Department
Imperial College London
p.dragotti@imperial.ac.uk

Course Motivation and Aim

- **Linear Algebra** is a branch of mathematics concerning vector spaces and linear mappings.

- Linear algebra is essential to model and understand many problems in engineering

- It is particularly useful
  - in statistics,
  - in optimization,
  - to design and understand filters and systems,
  - to solve problems related to approximation and estimation of signals,
  - In big-data related problems (e.g., DNN, graphs)

- **The main goal** of this module is to make you familiar with fundamental notions and methods in linear algebra while keeping a close eye on engineering applications

Imperial College London

Course Outline

- **Part 1**: Mathematical background
    - Vector Spaces, Inner Product, Norms
    - Orthogonality, Cauchy-Schwarz Inequality
    - Basis and Dimension
    - Finite-dimensional vector space and matrix notation
    - Linear mappings and linear transformation
    - Basics of Matrices (rank, determinant, trace, eigen-decompostion)
        - Examples: Markov Processes, Graphs, PageRank algorithm
    - Four Fundamental subspaces of a linear operator
    - Orthogonal Matrices, Gram-Schmidt orthogonalization process, QR factorization

Course Outline

- **Part 2**: How to solve $\boldsymbol{Ax} = \boldsymbol{b}$ (how to "*understand*" $\boldsymbol{Ax}=\boldsymbol{b}$)
  - Existence and Uniqueness
  - Left and Right Inverse
  - Full column-rank case
  - Projection theorem
  - Least-squares solution
  - Full row rank case
  - Minimum norm solution/least-squares with linear constraints
  - Pseudo-Inverse

- **Part 3**: SVD
  - SVD decomposition and connection with pseudo-inverse
  - Other applications of SVD: Total least-squares problems, approximation of rank deficient matrices, PCA

Material

Course Organization

- Lectures :18 hours
- Class and Revision problems: 3 hours
- Assessment: 100% examination in December
- Handouts in the form of pdf slides on BlackBoard
- Use “Ed Discussion” to post questions

Recommended Textbooks

- Introduction to Linear Algebra by Gilbert Strang.
- Mathematical Methods and Algorithms for Signal Processing by Todd K. Moon and Wynn C. Stirling (chapters 2-7)

Why Vector Spaces?

Systems of linear equations: Consider a system of 2 equations with 2 unknowns

$$2x - y = 0$$

$$-x + 2y = 3$$

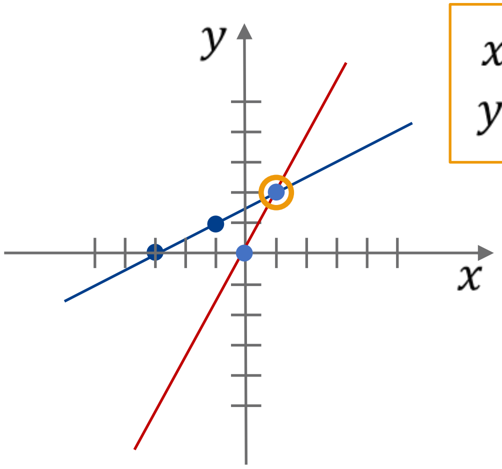
If we place the unknowns in a column vector, we can obtain the so called **matrix form** of the above system as follows.

$$\underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{\text{Coefficient Matrix}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{Vector of unknowns}} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Systems of linear equations: Row formulation

Consider the previous system of two equations.

$$\boxed{2x - y = 0}$$
$$\boxed{-x + 2y = 3}$$



$$\boxed{\begin{aligned} x &= 1 \\ y &= 2 \end{aligned}}$$

- Each equation represents a straight line in the 2D plane.
- The solution of the system is a point of the 2D plane that lies on both straight lines; therefore, it is their intersection.
- In that case, where the system is depicted as a set of equations placed one after another, we have the so called **Row Formulation** of the system.

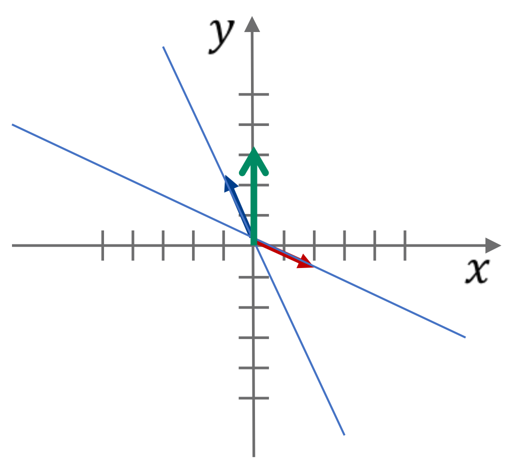
Systems of linear equations: Column formulation

- Have a look at the representation below:

$$2x - y = 0$$
$$-x + 2y = 3$$

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

- The weights of each unknown are placed jointly in a column vector.



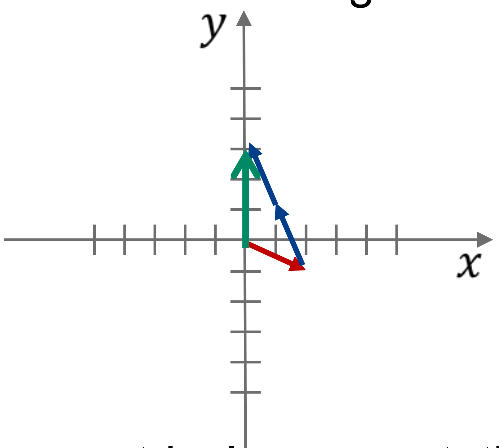
- The solution to the system of equations is **<u>that</u>** linear combination of the two column vectors that yields the vector on the right hand side.
- The above type of depiction is called **Column Formulation**.

Systems of linear equations: Column Formulation cont.

$$2x - y = 0$$
$$-x + 2y = 3$$

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

- The solution to the system of equations is the linear combination of the two vectors above that yields the vector on the right hand side.



- You can see in the figure a geometrical representation of the solution.

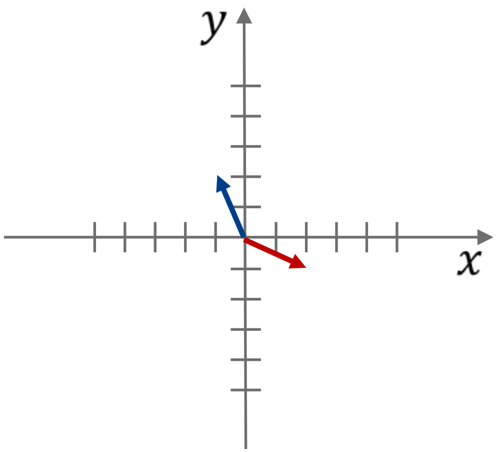
$$1 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \qquad x = 1, y = 2$$

Systems of linear equations: Column Formulation cont.

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

$$2x - y = 0$$
$$-x + 2y = 3$$

- What does the collection of ALL combinations of columns represents geometrically?



- All possible linear combinations of the columns form (span) the entire 2D plane!

Systems of linear equations: Matrix Formulation

- In matrix formulation we abandon scalar variables and numbers. Every entity is part of either a matrix or a vector as follows:

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

- In a real life scenario we have lots of equations and lots of unknowns. If we assume that we have m equations and n unknowns then we can depict the matrix formulation as:

$$\boldsymbol{Ax} = \boldsymbol{b}$$

where $\boldsymbol{A}$ is a matrix of size $m \times n$, $\boldsymbol{x}$ is a column vector of size $n \times 1$ and $\boldsymbol{b}$ is a column vector of size $m \times 1$.

Imperial College London

Systems of linear equations: Let's consider a higher order 3x3

Let us consider the row formulation of a system of 3 equations with 3 unknowns:

$$2x - y = 0$$

$$-x + 2y - z = -1$$

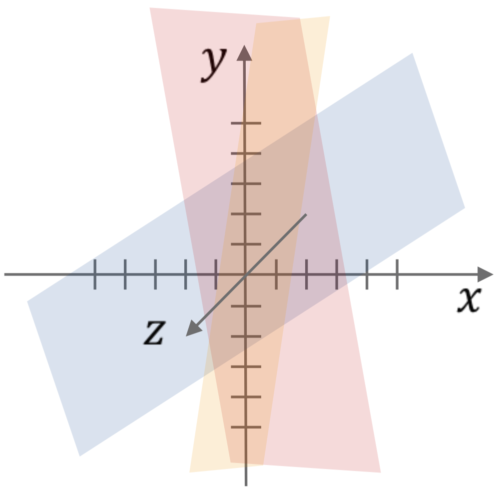
$$-3y + 4z = 4$$

Matrix Form: $Ax = b$

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

In the row formulation:

- Each row represents a plane on the 3D space.
- The solutions to the system of equations is the point where the 3 planes meet.
- As you can see the row formulation becomes harder to visualize for multi dimensional spaces!



Imperial College London

Systems of linear equations 3x3 cont.

- Let us now consider the column formulation of the previous system:

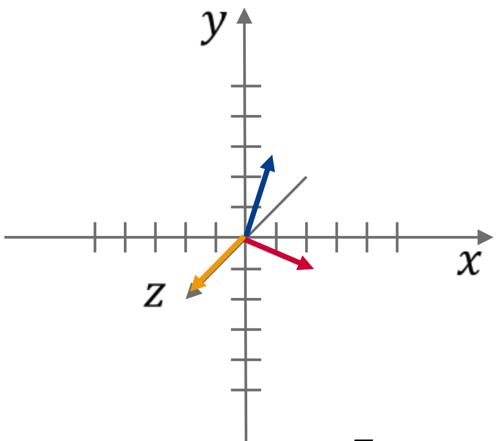
$$\begin{aligned} 2x - y \phantom{{}-z} &= 0 \\ -x + 2y - z &= -1 \\ -3y + 4z &= 4 \end{aligned}$$

Matrix Form: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

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

- Column formulation:

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



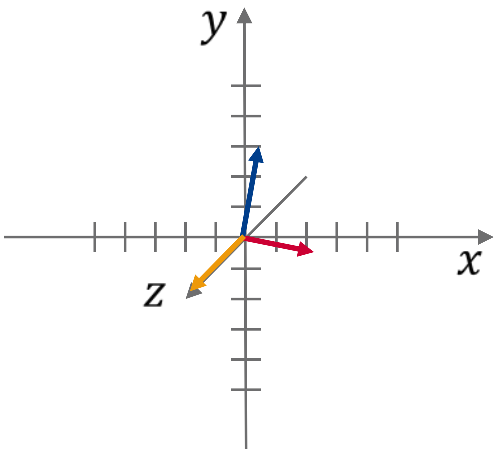
- The solution is **<u>that</u>** combination of column vectors which yield the right hand side. For the above system this is $[0 \quad 0 \quad 1]^T$.

Systems of linear equations: Is there always a solution?

- The solutions of a system of three equations with three unknowns lies inside the 3D space.
- Can you imagine a scenario for which there is no unique solution to the system?
- What if all three vectors lie on the same plane?
- Then there would not be a solution for every b.
- We will see later that in that case the matrix $\boldsymbol{A}$ would not be what is called **invertible**, since at least one of its column would be a linear combination of the other two.

Matrix Form: $A\,x = b$

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



We have had a first appreciation of matrix/vector formulation, but before we jump into the "*vector space*" world, let's revisit a practical algorithm to solve systems of linear equations.

Solving a system of linear equations using Gaussian Elimination (GE)

- A widely used method for solving a system of linear equations is the so called **Gaussian Elimination** (known also as **Row Reduction**).
- Consider the following system of 3 equations and 3 unknowns.

[2]-3[1]

$$\begin{aligned} x + 2y + z &= 2 \\ 3x + 8y + z &= 12 \\ 4y + z &= 2 \end{aligned}$$

- If we multiply the first row by 3 and subtract it from the second row we can eliminate x from the second row. We can use the notation $[2] - 3[1]$.

[3]-2[2]

$$\begin{aligned} x + 2y + z &= 2 \\ 2y - 2z &= 6 \\ 4y + z &= 2 \end{aligned} \qquad\qquad \begin{aligned} x + 2y + z &= \;\; 2 \\ 2y - 2z &= \;\; 6 \\ 5z &= -10 \end{aligned}$$

- Second step is to multiply the second row by 2 and subtract it from the third row, so we eliminate y from the third equation.

Imperial College London

Elimination and Back-substitution

- The solution to the system of linear equations after Gaussian Elimination, can be found by simply applying back-substitution.

$$\begin{aligned} x + 2y + z &= 2 \\ 2y - 2z &= 6 \\ 5z &= -10 \end{aligned} \qquad z = -2 \qquad \begin{aligned} y &= 1 \\ z &= -2 \end{aligned} \qquad \begin{aligned} x &= 2 \\ y &= 1 \\ z &= -2 \end{aligned}$$

- We can solve the equations in reverse order because the system after elimination is triangular.

Solving a system of linear equations using GE cont.

- So far, we have produced an equivalent representation of the system of equations.

$$\begin{aligned} x + 2y + z &= 2 \\ [2]-3[1] \quad 3x + 8y + z &= 12 \\ 4y + z &= 2 \end{aligned} \qquad \begin{aligned} x + 2y + z &= 2 \\ 2y - 2z &= 6 \\ [3]-2[2] \quad 4y + z &= 2 \end{aligned} \qquad \begin{aligned} x + 2y + z &= 2 \\ 2y - 2z &= 6 \\ 5z &= -10 \end{aligned}$$

- Consider how the matrix $\boldsymbol{A}$ is transformed.

$$[2]-3[1] \quad \begin{matrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{matrix} \qquad [3]-2[2] \quad \begin{matrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{matrix} \qquad \begin{matrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{2} & -2 \\ 0 & 0 & \boxed{5} \end{matrix}$$

Matrix $\boldsymbol{A}$ Matrix $\boldsymbol{u}$

pivots

- The **upper triangular matrix** we have produced is called $\boldsymbol{u}$, and the elements in the diagonal are called **pivots**.

Solving a system of linear equations using GE cont.

- So far, we have produced an equivalent representation of the system of equations.

$$\begin{array}{rr} & x + 2y + z = 2 \\ [2]-3[1] & 3x + 8y + z = 12 \\ & 4y + z = 2 \end{array} \qquad \begin{array}{rr} & x + 2y + z = 2 \\ & 2y - 2z = 6 \\ [3]-2[2] & 4y + z = 2 \end{array} \qquad \begin{array}{r} x + 2y + z = \ 2 \\ 2y - 2z = \ 6 \\ 5z = -10 \end{array}$$

- Similarly, the column b becomes:

$$\begin{array}{rr} & 2 \\ [2]-3[1] & 12 \\ & 2 \end{array} \quad \textbf{Matrix } \boldsymbol{b} \qquad \begin{array}{rr} & 2 \\ & 6 \\ [3]-2[2] & 2 \end{array} \qquad \begin{array}{r} 2 \\ 6 \\ -10 \end{array}$$

- It is often convenient to operate on the augmented matrix $[A|b]$.

$$\begin{array}{rrrrr} & 1 & 2 & 1 & 2 \\ [2]-3[1] & 3 & 8 & 1 & 12 \\ & 0 & 4 & 1 & 2 \end{array} \qquad \begin{array}{rrrrr} & 1 & 2 & 1 & 2 \\ & 0 & 2 & -2 & 6 \\ [3]-2[2] & 0 & 4 & 1 & 2 \end{array} \qquad \begin{array}{rrrr} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array}$$

Augmented matrix

Elimination viewed as matrix multiplication

- Let us consider again the steps of elimination.

$$\begin{aligned} x+2y+z&=2\\ [2]-3[1] \quad 3x+8y+z&=12\\ 4y+z&=2 \end{aligned} \qquad \begin{aligned} x+2y+z&=2\\ 2y-2z&=6\\ [3]-2[2] \quad 4y+z&=2 \end{aligned} \qquad \begin{aligned} x+2y+z&=2\\ 2y-2z&=6\\ 5z&=-10 \end{aligned}$$

- Observe how the matrix A is transformed in each step:

$$[2]-3[1] \quad \begin{matrix} 1 & 2 & 1\\ 3 & 8 & 1\\ 0 & 4 & 1 \end{matrix} \qquad [3]-2[2] \quad \begin{matrix} 1 & 2 & 1\\ 0 & 2 & -2\\ 0 & 4 & 1 \end{matrix} \qquad \begin{matrix} 1 & 2 & 1\\ 0 & 2 & -2\\ 0 & 0 & 5 \end{matrix}$$

Matrix A

- Each step where a substitution of the form $[i]+c*[j]$ takes place, is equivalent to multiplying the current matrix with an identity matrix whose $[i,j]$ element has been replaced by c. This matrix is denoted with $\boldsymbol{E}_{ij}$.

$$[2]-3[1] \quad \underbrace{\begin{bmatrix} 1 & 0 & 0\\ \boxed{-3} & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}}_{E_{21}} \begin{bmatrix} 1 & 2 & 1\\ 3 & 8 & 1\\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1\\ 0 & 2 & -2\\ 0 & 4 & 1 \end{bmatrix}$$

Elimination viewed as matrix multiplication cont.

- Let us consider again the steps of elimination.

$$\begin{array}{rl} & x+2y+z=2 \\ [2]-3[1] & 3x+8y+z=12 \\ & 4y+z=2 \end{array} \qquad \begin{array}{rl} & x+2y+z=2 \\ & 2y-2z=6 \\ [3]-2[2] & 4y+z=2 \end{array} \qquad \begin{array}{r} x+2y+z=\ \ 2 \\ 2y-2z=\ \ 6 \\ 5z=-10 \end{array}$$

- Observe how the matrix A is transformed in each step:

$$\begin{array}{r} \\ [2]-3[1] \\ \\ \end{array} \begin{matrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{matrix} \qquad \begin{array}{r} \\ \\ [3]-2[2] \end{array} \begin{matrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{matrix} \qquad \begin{matrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{matrix}$$

Matrix A

- The second step of elimination can be viewed as the following matrix multiplication:

$$[3]-2[2] \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \boxed{-2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

Elimination viewed as matrix multiplication cont.

- Therefore, the elimination can be expressed in matrix form as:

$$E_{32}(E_{21}A) = u$$

- Brackets can be obviously dropped, therefore:

$$E_{32}E_{21}A = u$$

- It is not hard to prove that the inverse of each elimination matrix is obtained by replacing its non-zero off-diagonal element with its reversed sign value.

$$E_{21}^{-1} \quad \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \underset{E_{21}}{\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underset{E_{32}^{-1}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}} \underset{E_{32}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

LU Decomposition

- We showed that the entire process of elimination can be expressed as a sequence of matrix multiplications.

- The original system matrix A is multiplied by a sequence of matrices which have a simple form and their inverses have a simple form too. In the previous example we have:

$$E_{32}E_{21}A = u$$

- In the above equation, if we sequentially multiply both sides from the left with the inverses of the individual elimination matrices we obtain:

$$A = E_{21}^{-1}E_{32}^{-1}u$$

- Therefore, matrix $\boldsymbol{A}$ can be decomposed as: $\boldsymbol{A = LU}$

$$L = E_{21}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

- It can be proven that $\boldsymbol{L}$ is a **lower triangular** matrix.
- **The above formulation is called $\boldsymbol{LU}$ decomposition.**

LU Decomposition in the general case

- We have seen previously that elimination can be viewed as a multiplication of a series of elimination matrices, e.g. in the 3×3 case we have the general form:

$$E_{32}E_{31}E_{21}A = u$$

- By multiplying with the inverses of the elimination matrices in reverse order we get:

$$A = {E_{21}}^{-1}{E_{31}}^{-1}{E_{32}}^{-1}u$$

- The product of the inverses of the elimination matrices is:

$$L = {E_{21}}^{-1}{E_{31}}^{-1}{E_{32}}^{-1}$$

- Matrix L has the nice property that its elements are the multipliers used in elimination.

$$A = L\,u$$

LU Decomposition with row exchanges. Permutation.

- Often in order to create the upper triangular matrix u through elimination we must reorder the rows of matrix A first
- In the general case where row exchanges are required, for any invertible matrix A, we have:

$$PA = Lu$$

- P is a **permutation** matrix. This arises from the identity matrix if we reorder the rows.
- A permutation matrix encodes row exchanges in Gaussian elimination.
- Row exchanges are required when we have a zero in a pivot position.
- For example the following permutation matrix exchanges rows 1 and 2 to get a non zero in the first pivot position

$$\overset{P_{12}}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} \boxed{0} & 1 & 0 \\ 3 & 1 & 2 \\ 2 & 5 & 3 \end{bmatrix} = \begin{bmatrix} \boxed{3} & 1 & 2 \\ 0 & 1 & 0 \\ 2 & 5 & 3 \end{bmatrix}$$

Vector Spaces

The notions of vector spaces, inner product etc. are necessary to create a common framework to handle vectors in dimension ≤ 3 (for which we can use geometry) and vectors in dimension $n > 3$.

Definition 1: By a **vector space** we mean a collection of objects known as vectors that form a nonempty set E with two operations an addition ($E \times E \to E$) and a scalar multiplication that satisfy the following properties:

- Commutativity: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$;
- Associativity $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$;
- Distributivity $(\alpha + \beta)\, \boldsymbol{x} = \alpha\boldsymbol{x} + \beta\boldsymbol{x}$ and $\alpha(\boldsymbol{x} + \boldsymbol{y}) = \alpha\, \boldsymbol{x} + \alpha\, \boldsymbol{y}$;
- For every $\boldsymbol{x}, \boldsymbol{y} \in E$ there exists $\boldsymbol{z} \in E$ such that $\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}$;
- $\alpha(\beta\, \mathbf{x}) = (\alpha\beta)\boldsymbol{x}$;
- $1\boldsymbol{x} = \boldsymbol{x}$.

Inner Product

Definition 2: Let E be a complex vector space. A mapping $\langle .,. \rangle$ from $E \times E \to \mathbb{C}$ is called an inner product in E if for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in E$ and $\alpha, \beta \in \mathbb{C}$ the following conditions are satisfied:

- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$
- $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$
- $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ implies $\boldsymbol{x} = 0$

Inner Product and Norm

It is always useful to be able to determine the size of an object, in the case of vector spaces this is achieved using the notion of 'norm'.

Definition 3: A **norm** on a vector space E over $\mathbb{C}$ (or $\mathbb{R}$) is a real valued function with the following properties for any $\boldsymbol{x}, \boldsymbol{y} \in E$ and $\alpha \in \mathbb{C}$:

- Positive definiteness $||\mathbf{x}|| \geq 0$, and $||\mathbf{x}|| = 0$ if and only if $\boldsymbol{x} = 0$.
- Homogeneity Property: $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$.
- Triangle inequality: $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$

Given the definition of inner product, we define the **induced norm** of a vector $\boldsymbol{x}$ as:

$$||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Examples of Vector Spaces

Example 1: n-dimensional vectors in $\mathbb{R}^n$or $\mathbb{C}^n$. A **vector** is a set of scalars placed jointly in a vertical fashion (column vector) .

- A column vector of size n is denoted as $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$
- The Hermitian transpose of $\boldsymbol{x}$ is a row vector of size n denoted as $\boldsymbol{x}^H = (x_1^* \quad \dots \quad x_n^*)$
- Addition and product with a scalar are easily defined, while the inner product is:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^H \boldsymbol{x} = \sum_{i=1}^{n} x_i y_i^*$$

- Consequently, the induced norm is:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

On the notion of norm

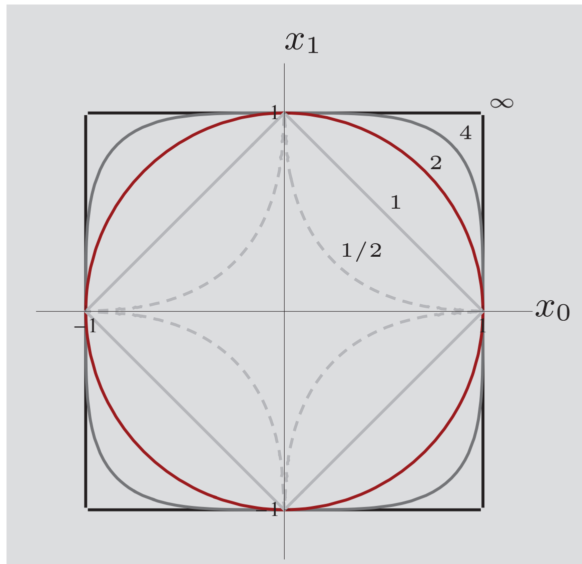
The induced norm inspires a more general definition of norm which is often used in modern signal processing:

- Induced norm:

$$\|\boldsymbol{x}\|_2 = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$$

- More general norm: $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

On the Notion of the Norm (cont'd)



Sets of unit-norm vectors for different p norms. Note that $\|x\|_\infty = \sup_{|} x_n|$ and that $p = 1/2$ is not a norm. The $\|x\|_0$ corresponds to the number of non-zero entries in x and is not a norm.

Figure taken from 'Foundations of Signal Processing', Vetterli, Kovacevic and Goyal, Springer 2014.

Examples of Vector Spaces

Example 2: the space of matrices of size $m \times n$, that is, the space with elements $\boldsymbol{A} \in \mathbb{C}^{m \times n}$

- We tend to use matrices to model linear equation, but they are actually more complex objects

- The inner product of two matrices ***A***,***B*** is the element-by-element product: $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j} a_{i,j} b_{i,j}^*$

- This leads to the induced norm (**Frobenius norm**): $\|\boldsymbol{A}\|_{\boldsymbol{F}} = \sqrt{\sum |a_{i,j}|^2}$

- More general norms: $\|\boldsymbol{A}\|_p = \sup \frac{\|Ax\|_p}{\|x\|_p}$

Cauchy-Schwarz inequality

- Cauchy-Schwarz inequality: $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$
- The Cauchy-Schwarz inequality allows us to estimate the similarity of two vectors and to introduce the idea of direction.
- We have the angle between two vectors: $\cos\theta = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$
- So two vectors are '*maximally*' similar when one is the rescaled version of the other
- They are '*maximally*' dissimilar when $\theta = \frac{\pi}{2}$, that is, when $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$
- If the inner product of two vectors is zero, the vectors are called **orthogonal**.
- Additionally, if their magnitudes are 1 they are called **orthonormal**.
- Remember that this applies to any inner-product vector space.

Proof of Cauchy-Schwarz inequality

Cauchy-Schwarz inequality: $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$

Sketch of the proof:

- We have:

$$0 \leq \|\boldsymbol{x} - \alpha\, \boldsymbol{y}\|^2 = \langle \boldsymbol{x} - \alpha\, \boldsymbol{y}, \boldsymbol{x} - \alpha\, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 - \langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle - \langle \alpha \boldsymbol{y}, \boldsymbol{x} \rangle + |\alpha|^2 \|\boldsymbol{y}\|^2$$

- Now assuming: $\alpha = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}$ and using that fact that if $c = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ then $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = c^*$, we obtain

$$0 \leq \|\boldsymbol{x} - \alpha\, \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 - \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^2}$$

which yields the Cauchy-Schwarz inequality

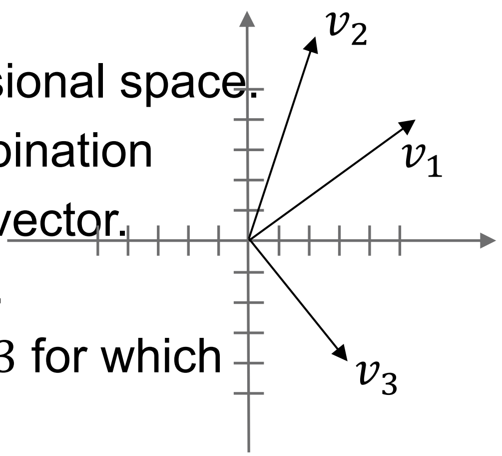
Linear Combination and Linear Independence

- Let $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, .. \boldsymbol{v_n}$ be vectors in a vector space of dimension n. Then for $c_i \in \mathbb{R}$ the linear combination $\boldsymbol{x} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + c_3\boldsymbol{v_3} + .. c_n\boldsymbol{v_n}$ is in the same vector space and can be expressed in matrix/vector form:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

- The vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, .. \boldsymbol{v_n}$ are **independent** if no linear combination of them gives the zero vector (except the zero combination, all $c_i = 0$)

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + c_3\boldsymbol{v_3} + .. c_n\boldsymbol{v_n} \neq 0$$



- Take two non-zero, non-parallel vectors in the two-dimensional space.
- They are independent because there isn't any linear combination with non-zero coefficients of them that can give the zero vector.
- Now consider three vectors in the two-dimensional space. These are dependent. That means there exist x_i, $i = 1,2,3$ for which

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = 0$$

Span

Definition 5: Let T be a set of vectors in a vector space E. The set V of all vectors 'reachable' by taking linear combinations of vectors in T is the **span** of the vectors. This is denoted by $V = span\{T\}$.

This means that for any $\boldsymbol{x} \in V$ we can write: $\boldsymbol{x} = \sum_i c_i \boldsymbol{v}_i$

Note that by construction V is a vector space and since $V \subseteq E$ we call it a sub-space.

Example: The span of two linearly independent vectors in $\mathbb{R}^3$ forms a plane which is a subspace of $\mathbb{R}^3$.

Is the following plane a subspace as well?

$$V = \{x \in \mathbb{R}^3 \colon x_1 + x_2 + x_3 = 1\}$$

Basis and Dimension

Definition 6: Let E be a vector space and let T be a set of vectors from E such that $span\{T\} = E$. If the vectors in T are linearly independent then they form a **basis** for E. The cardinality of T is the **dimension** of E.

Example: Consider either the space $\mathbb{R}^n$ or $\mathbb{C}^n$, the set of vectors $\boldsymbol{e}_i = (0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0)^{\mathrm{T}}$ *i=1,2,…,n,* where only the *i*th element of $\boldsymbol{e}_i$ is one, is a basis for $\mathbb{R}^n$ or $\mathbb{C}^n$ and is known as the **standard** or **canonical** basis

When a set of vector $\boldsymbol{v}_i$ $i = 1,2, \dots, n$ forms a basis and they are also orthogonal (i.e., $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ when $i \neq j$) then we have an **orthogonal basis**. Moreover, if $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = 1$, $i = 1,2, \dots, n$ the basis is **orthonormal**.

Exercise: Consider the set $S = \{x = (1,2,\gamma)^T\colon \gamma \in \mathbb{R}\} \subset \mathbb{R}^3$. What is the dimension of *span*{*S*}?

Answer: $span\{S\} = \{x = (\alpha, 2\alpha, \beta)^T\colon \alpha, \beta \in \mathbb{R}\} = span\{(1,2,0)^T, (0,0,1)^T\}$ so S has dimension 2.

Orthogonal (unitary) Matrices

- Consider a set of vector $\boldsymbol{v}_i$ $i = 1,2,\ldots,n$ which forms an **orthonormal** basis and create a matrix $\boldsymbol{A}$ from stacking them one after the other:

$$\boldsymbol{A} = \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{pmatrix}$$

- Clearly $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{I}$, where $\boldsymbol{A}^H$ is the Hermitian transpose of $\boldsymbol{A}$ and $\boldsymbol{I}$ is the identity matrix. This means that $\boldsymbol{A}^H = \boldsymbol{A}^{-1}$.

- This type of matrices are called orthogonal (unitary) matrices and **they preserve the induced norm**: Assume $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ then

$$||\boldsymbol{y}||^2 = \boldsymbol{y}^{\boldsymbol{H}}\boldsymbol{y} = \boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = ||\boldsymbol{x}||^2$$

If the set of vector $\boldsymbol{v}_i$ $i = 1,2,\ldots,n$ is not orthonormal, the process to make them orthogonal is known as Gram-Schmidt process.

Examples of Orthogonal Matrices

- The **rotation matrix** of size 2×2 is defined as:
  $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

- **Permutation** matrices reorder the rows of identity matrices. Examples are:
  $Q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ and $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. $Q^T = Q^{-1}$ in both cases.

Problems:

- The columns of Q are orthogonal (straightforward).
- The columns of Q are unit vectors (straightforward).
- Explain the effect that the permutation matrices have on an arbitrary vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ when they multiply the vector from the left.
- Obviously the matrices cause re-ordering of the elements of these vectors

Orthogonal Subspaces

We have learned the notion of orthogonality for two vectors. We now want to extend it to subspaces. In that case we require that *every* vector of one subspace is orthogonal to *every* vector in the other.

Definition 7: Let E be a vector space and let V and W be subspaces of E. V and W are orthogonal if every vector $\boldsymbol{v} \in V$ is orthogonal to every vector $\boldsymbol{w} \in W$: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$.

Definition 8: For a subspace $V \subset E$, the subspace of all vectors orthogonal to every vector in V is called the orthogonal complement to V and is denoted as $V^{\perp}$ $(E = V \oplus V^{\perp})$.

Linear Mappings and Matrices

- We are used to think of matrices (and vectors) as a good way to express linear equations in compact form.
- Instead, now that we have gained insights into vector spaces, we want to give matrices a broader and more 'geometrical' interpretation.
- Consider two vector spaces X and Y of dimension n and m respectively, a *map* or a *transformation* is a function that assigns to each element of X (called the domain), a unique element in Y (called the range) and is denoted as $X \Rightarrow Y$.
- If the mapping is linear, it can be represented using a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$

Imperial College London

Linear Mapping: Examples and Exercises

- Assume $X, Y \in \mathbb{R}^2$, can you work out what the following mapping is doing to the vectors in X?

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} \quad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

- Can you describe a 2-D rotation by 2θ using the product of two matrices?
- Can you invert a 2-D rotation by θ with a proper new linear mapping?

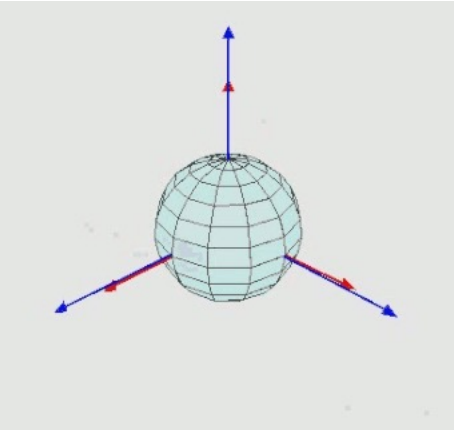
Imperial College London

Linear Mapping: Examples and Exercises

- Assume $X, Y \in \mathbb{R}^2$, can you work out what the following mapping is doing to the vectors in X?

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} \quad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

- Can you describe with a matrix a rigid rotation in 3-D?



Note: Animation taken from Wikipedia “Euler Angles”

Linear Mapping: Examples and Exercises (cont'd)

- Many important operations in signal processing can be described using matrices.
  - For example given a finite duration discrete-time signal $\boldsymbol{x} = (x_1, x_2, \dots x_n)^T$ you want to find its time-reversed version $\boldsymbol{y} = (x_n, \dots, x_2, x_1)^T$ can you describe this mapping with a proper matrix?
  - Assume a discrete-time signal $\boldsymbol{x} = (x_1, x_2, \dots x_n)^T$ has size n, you want to truncate it by removing the last $m - n$ terms with $m < n$. This operation can be described with a 'fat' $m \times n$ matrix with ones along the diagonal
  - Can you describe the 'zero-padding' operation using matrices?
  - Can you describe the permutation operation using the *permutation* matrix?
  - Can you describe the differentiation of polynomials of degree d using a $d \times d$ matrix?

Imperial College London

Circulant Shift

- Assume you want to shift the vector $\boldsymbol{x} = (x_1, x_2, \dots x_n)^T$ by one unit.
- You may impose that the first entry moves to the end. This is equivalent to treating $\boldsymbol{x}$ as a periodic sequence of period n
- In matrix/vector form with $n = 4$ we have:
- $\boldsymbol{Px} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ x_1 \end{bmatrix}$
- If we multiply x by $\boldsymbol{P}^2$ we shift by 2 → $= (x_3, x_4, x_{1,} x_2)^T$
- We shift by 3 if we multiply the original vector by $\boldsymbol{P}^3$

Circulant matrix

- The matrix $\boldsymbol{C}$ given by the linear combination of $\boldsymbol{I}, \boldsymbol{P}, \boldsymbol{P}^2, \boldsymbol{P}^3$ is called a circulant matrix and is a very important matrix to describe filtering.

- $\boldsymbol{C} = c_{\boldsymbol{0}}\boldsymbol{I} + c_{\boldsymbol{1}}\boldsymbol{P} + c_2\boldsymbol{P}^2 + c_3\boldsymbol{P}^3 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 \\ c_3 & c_0 & c_1 & c_2 \\ c_2 & c_3 & c_0 & c_1 \\ c_1 & c_2 & c_3 & c_0 \end{bmatrix}$

Discrete-time Convolution

- Given a discrete-time (real-valued) filter with unit impulse response h_k and an input sequence x_n, the output y_n is related to the input through the convolution formula

$$y_n = \sum_k h_k x_{n-k}$$

- Assume the impulse response of the filter has finite duration (FIR filter), e.g., $h_k \neq 0$ for $k = 0,1, \dots, K-1$ and that the input sequence has also finite duration $n > K$

- Under these conditions, convolution is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$:

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

Discrete-time Convolution and Toeplitz matrices

- Under these conditions, convolution is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$:

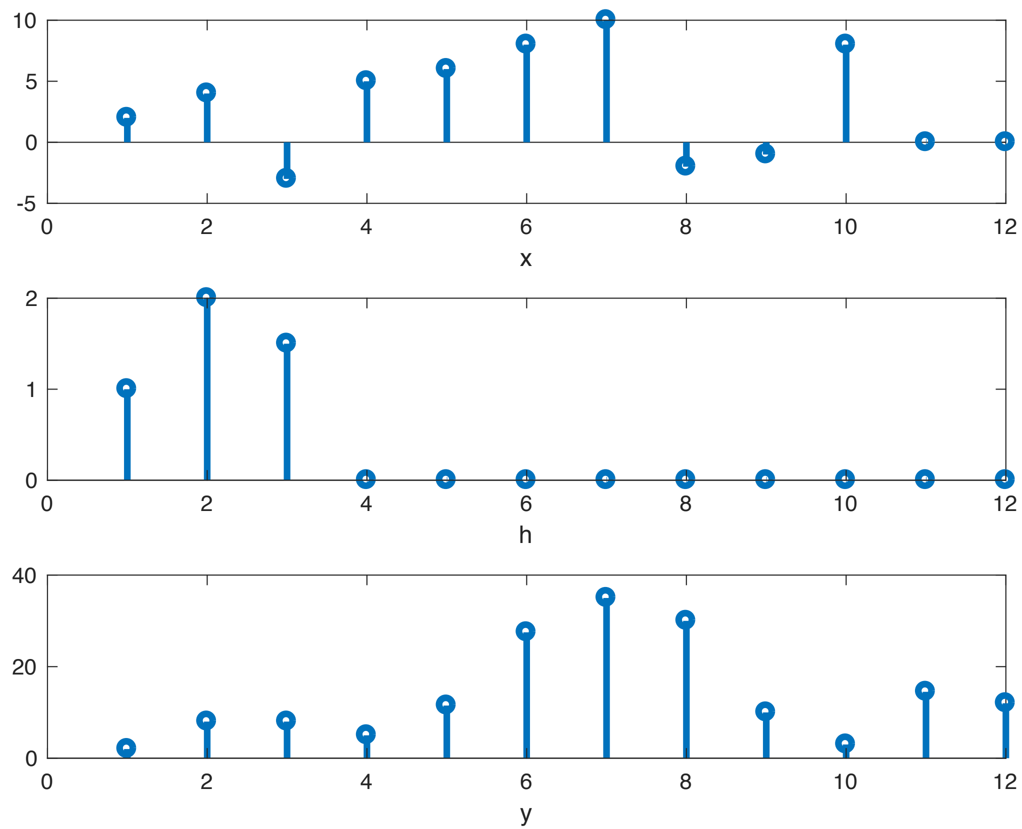
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

- $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has the same entry along the diagonals and is called a '**Toeplitz**' matrix

$$\boldsymbol{A} = \begin{pmatrix} h_0 & 0 & \cdots & \cdots & \cdots & 0 \\ h_1 & h_0 & \cdots & \cdots & \cdots & \vdots \\ \vdots & h_1 & \ddots & \ddots & \ddots & 0 \\ h_{K-1} & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & h_{K-1} & \vdots & \vdots & \vdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & h_1 & h_0 \\ 0 & \ddots & \ddots & \ddots & h_2 & h_1 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & h_{K-1} \end{pmatrix}$$

Discrete-time Convolution and Circulant Matrices

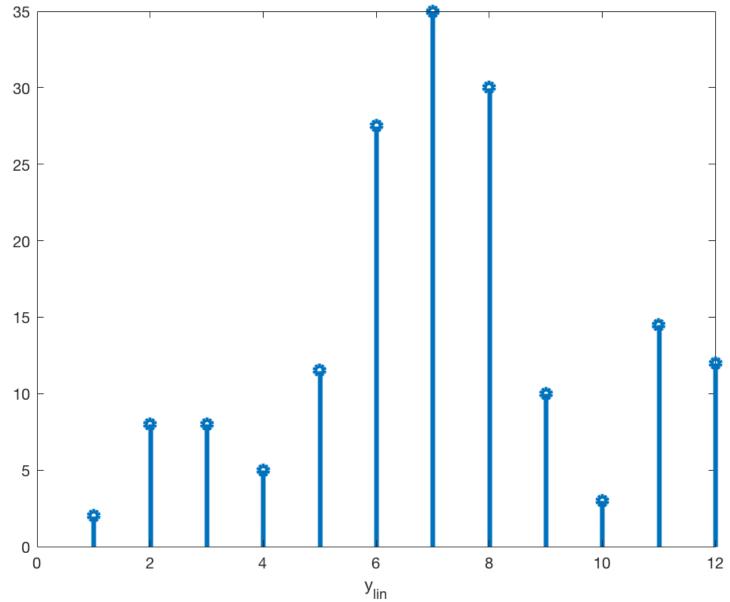
- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $m = n + K - 1$ models the ***linear*** convolution.

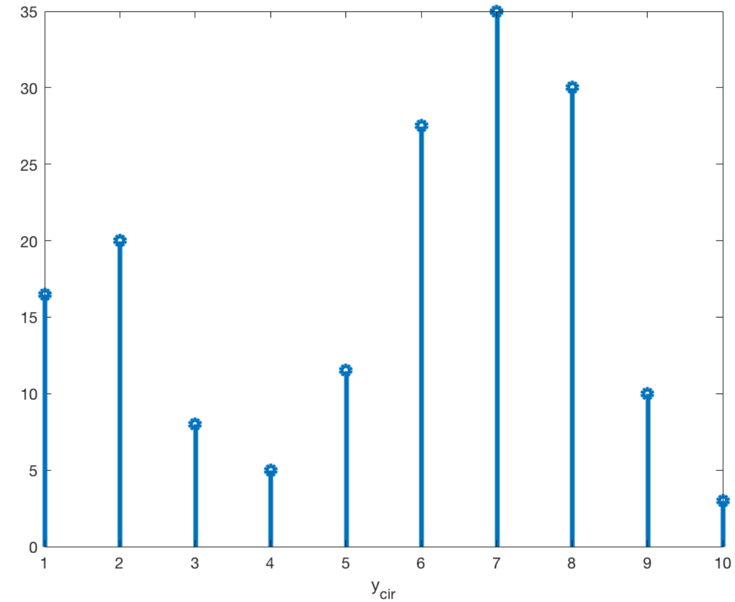


Imperial College London

Discrete-time Convolution and Circulant Matrices

- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $m = n + K - 1$ models the ***linear*** convolution.
- If we want $\boldsymbol{A}$ to be square, we may assume that $\boldsymbol{x} \in \mathbb{R}^{n}$ is periodic outside the interval of observation. In this case $\boldsymbol{A}$ is '**circulant'** matrix and models the ***circular*** convolution.





Discrete-time Convolution: Example

- **Example**: assume you want to implement a discrete-time version of the derivative $y(t) = \frac{dx(t)}{dt}$. The natural way to do it is by taking finite differences of the incoming sequence: $y_n = x_n - x_{n-1}$. This is achieved by filtering with $h_k = [1;\ -1]$:

$$y_n = \sum_k h_k x_{n-k}$$

In the case of linear convolution, the linear mapping is described by the $(n+1)\times n$ matrix:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & \cdots & \cdots & \cdots & \vdots \\ \vdots & -1 & \ddots & \ddots & \ddots & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \vdots & \vdots & \vdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}$$

Discrete-time Convolution: Example (cont'd)

In the case of circular convolution, the linear mapping is described by the $n \times n$ matrix:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & -1 \\ -1 & 1 & \ddots & \ddots & 0 \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & -1 & 1 \end{pmatrix}$$

Discrete-time Fourier Transform

The discrete Fourier transform (DFT) is a linear mapping and can be described with a matrix.

Recall that, given the finite-length sequence: $x_0, \; x_1, \dots, x_{N-1}$, its DFT is given by:

$$X_r = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{-j2\pi\frac{rn}{N}} \qquad\qquad r = 0,1, \dots N-1$$

In matrix/vector form: $\boldsymbol{X} = \boldsymbol{F}^H \boldsymbol{x}$ where ***F*** is given by:

$$\boldsymbol{F} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{j\frac{2\pi}{N}} & e^{j\frac{4\pi}{N}} & \vdots & e^{j2\pi\frac{N-1}{N}} \\ 1 & e^{j\frac{4\pi}{N}} & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & e^{j2\pi\frac{(N-2)(N-2)}{N}} \\ 1 & e^{j2\pi\frac{N-1}{N}} & \cdots & e^{j2\pi\frac{(N-1)(N-2)}{N}} & e^{j2\pi\frac{(N-1)(N-1)}{N}} \end{pmatrix}$$

Discrete-time Fourier Transform

- We call $\boldsymbol{F}^H$ the DFT matrix (note the normalization term $\frac{1}{\sqrt{N}}$).
- Also note that $\boldsymbol{F}\boldsymbol{F}^H = \boldsymbol{I}$ so the discrete Fourier transform is a perfectly invertible linear mapping.
- This also shows that the DFT matrix is a ***unitary*** matrix (columns are orthonormal vectors)
- Examples: 2 points DFT:

$$\boldsymbol{F}^H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- 4 points DFT

$$\boldsymbol{F}^H = \frac{1}{\sqrt{4}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^9 \end{bmatrix} = \frac{1}{\sqrt{4}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \text{ with } \omega = e^{-\frac{j2\pi}{4}}$$

Range and Null Space of a Linear Transformation

- Consider the linear transformation $X \Rightarrow Y$ represented by

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

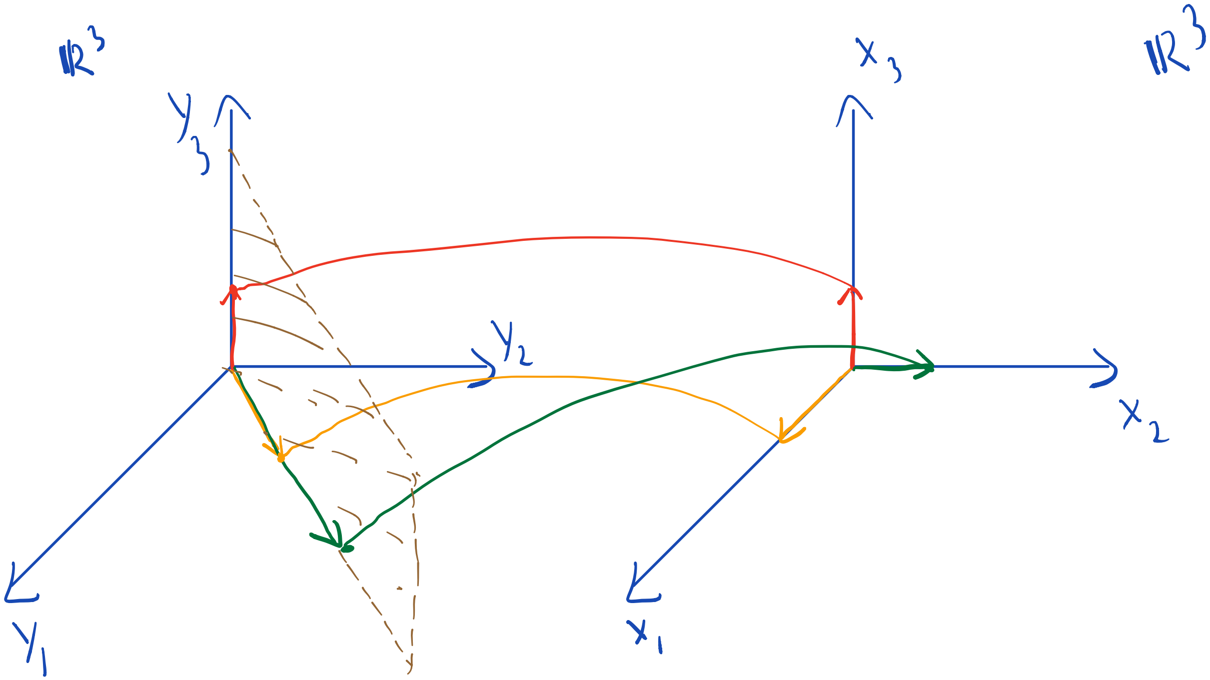
- Clearly, eventhough the mapping is from $\mathbb{R}^3$ to $\mathbb{R}^3$, the elements of X are 'collapsed' into a plane. This gives us immediately the feeling that this mapping is not invertible.

- The notion of range space and null space of a mapping makes this intuition more precise

Range and Null Space of a Linear Transformation, example

- Consider the linear transformation $X \Rightarrow Y$ represented by

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



Range and Null Space of a Linear Transformation

- Consider the linear transformation $X \Rightarrow Y$ represented by

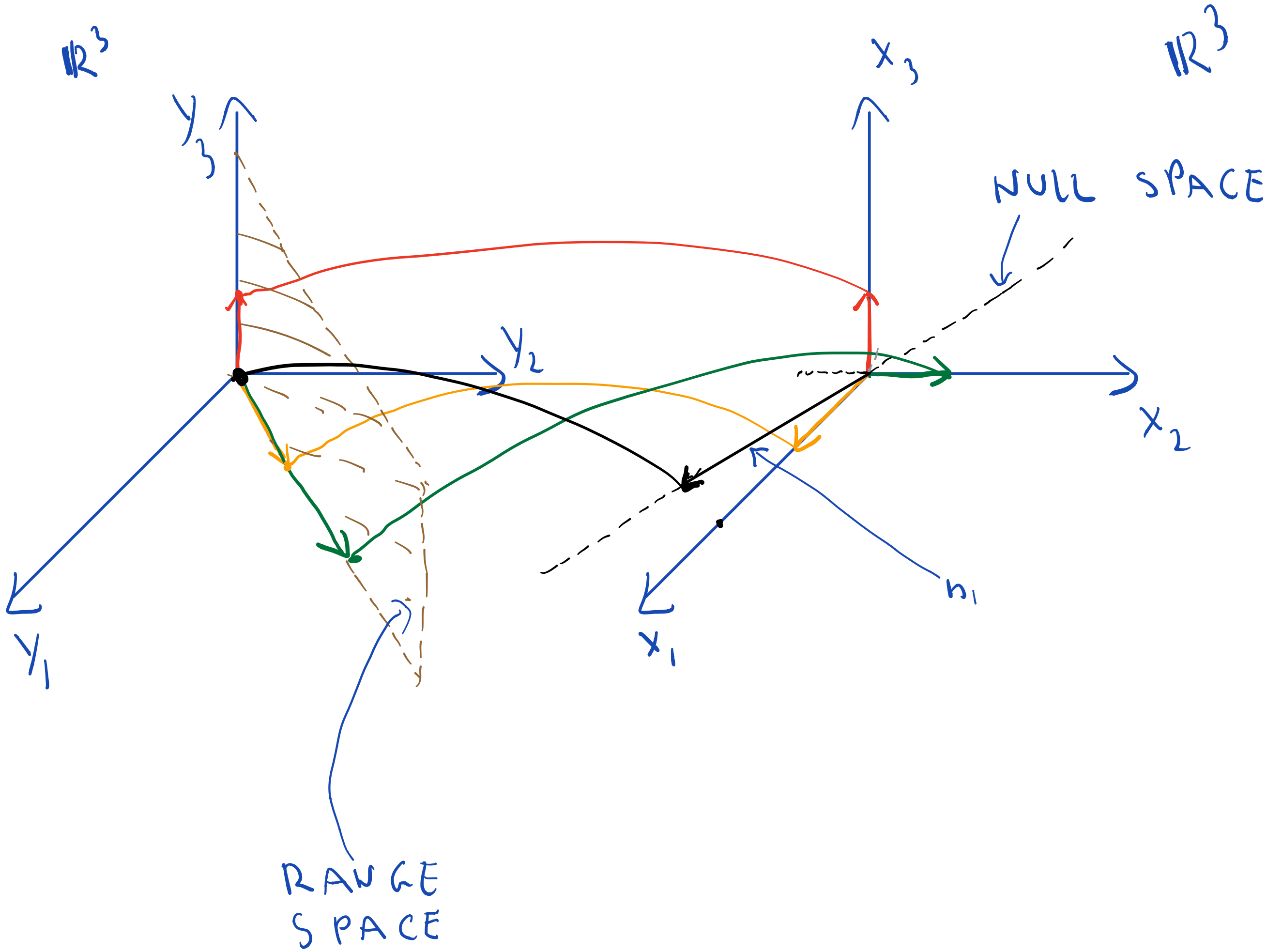
$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Preliminary observations

  - The columns of $\boldsymbol{A}$ are linearly dependent
  - The first and third column are the only independent columns
  - The plane where all the vectors $\boldsymbol{x}$ were collapsing to is span$\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\}$ and this is the range of the matrix

- We also note that if $\boldsymbol{x} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$ then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ (this specific $\boldsymbol{x}$ is in the null space of $\boldsymbol{A}$)

Range and Null Space



Imperial College London

Range and Null Space of a Linear Transformation

- **Definition 9**: Let A: $X \Rightarrow Y$ be a linear transformation. The **range space** $\mathcal{R}(A)$ is the set of values in Y that are reached from X by applying $\boldsymbol{A}$:

$$\Re(A) = \{\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathrm{X}\}$$

The **null space** $\mathcal{N}(A)$ is the set of values in X that are transformed to $\mathbf{0}$ by $\boldsymbol{A}$:

$$\mathcal{N}(A) = \{\boldsymbol{x} \in X : \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\}$$

- Let $\boldsymbol{A}$ be an $m \times n$ matrix which we regard as a linear operator and which we write as follows:

$$\boldsymbol{A} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \cdots \quad \boldsymbol{v}_n]$$

then a vector $\boldsymbol{x} \in \mathbb{R}^n$ is transformed as:

$$\boldsymbol{A}\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n$$

which is a linear combination of the columns of $\boldsymbol{A}$.

Range and Null Space of a Linear Transformation

- This means that the range can be expressed as

$$\Re(A) = span\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\}$$

  and corresponds to the column space of $\boldsymbol{A}$. Moreover, the number of linear independent columns of $\boldsymbol{A}$, which by definition is also the rank of $\boldsymbol{A}$, corresponds to the dimension of $\Re(A)$

- The null space of $\boldsymbol{A}$ is also a subspace and its dimension is n – rank($\boldsymbol{A}$).

- When the dimension of $\mathcal{N}(\boldsymbol{A}) > 0$ then the mapping is not invertible. Assume that $\boldsymbol{A}\boldsymbol{n}_1 = \boldsymbol{0}$ and that $\boldsymbol{y}_1 = \boldsymbol{A}\boldsymbol{x}_1$ then, given $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{n}_1$, we have that $\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{n}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{y}_1$. Therefore given $\boldsymbol{y}_1$ it is not possible to know whether it is due to $\boldsymbol{x}_1$ or $\boldsymbol{x}_2$

Range and Null Space of a Linear Transformation

- The null space of $\boldsymbol{A}$ is also a subspace and its dimension is $n - \text{rank}(\boldsymbol{A})$.

- Sketch of the Proof:
  - Remember that when we compute $\boldsymbol{A}\boldsymbol{x}$ we are effectively taking a linear combination of the columns of $\boldsymbol{A}$
  - Assume for the sake of argument that $rank(\boldsymbol{A}) = n - 2$, this means that given $n - 2$ columns of $\boldsymbol{A}$, the other two columns can be obtained with proper linear combinations of the linearly independent $n - 2$ columns.
  - Therefore there are two linearly independent vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ such that $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{0}$
  - Moreover any linear combination of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is also in the null space of $\boldsymbol{A}$, therefore that space has dimension 2.

- Assume that $\boldsymbol{A}$ models a linear mapping from $\mathbb{C}^n$ to $\mathbb{C}^n$ and that its null space is *non-trivial* (i.e., it has dimension larger than 0 or $rank(\boldsymbol{A}) < n$) then the mapping is not invertible or in other words the matrix $\boldsymbol{A}$ does not have an inverse.

- In this case we call $\boldsymbol{A}$ a **singular** matrix.

Computing the Nullspace (example)

- We will solve the system $\boldsymbol{Ax} = \mathbf{0}$ by elimination.

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

- During elimination the nullspace remains unchanged, since the solution to $\boldsymbol{Ax} = \mathbf{0}$ does not change by elimination.
- The first two steps of elimination yield the matrix below right.
- Note that we can't find a '*pivot*' in the second column, meaning that the second column is not independent (depends on the previous column).

$$\begin{array}{l} \\ [2]-2[1] \\ [3]-3[1] \end{array} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix}$$

Computing the Nullspace

- We continue the elimination in the third column.
- We also notice that the last column doesn't have a pivot and it also depends on the previous columns.

$$\begin{array}{l} [2]-2[1] \\ [3]-3[1] \end{array} \begin{matrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{matrix} \qquad [3]-[2] \quad \begin{matrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{matrix} \qquad \begin{matrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 0 & 0 \end{matrix}$$

- Therefore, in this case we only have 2 pivots, signifying the number of independent columns.
- The number of pivots is the rank of the matrix.
- Therefore, in this particular example $rank(A) = 2$.
- Note that in the rectangular **(non-square)** case the resulting matrix u is not really an upper triangular, but it is in the so-called **Echelon** (staircase) form.
- In order to identify the nullspace we need to describe the solutions of $\boldsymbol{Ax} = \mathbf{0}$.

Computing the Nullspace

- By applying elimination we obtained the matrix $\boldsymbol{u}$.

$$\boldsymbol{u} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- The matrix u contains two **pivot columns** shown in **red** and two **free columns** shown in **blue**.
- The free columns represent free variables, i.e., variables that we can assign any values to them.
- We obtain the null space of $\boldsymbol{Ax} = \boldsymbol{0}$, by solving the system $\boldsymbol{ux} = \boldsymbol{0}$.

Imperial College London

Computing the Nullspace

- In order to calculate the null space we need to solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.
- The system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is equivalent to $ux = 0$ which can be written as:

$$\boldsymbol{u}\boldsymbol{x} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \boldsymbol{0}$$

- Using row formulation we obtain:

$$x_1 + 2x_2 + 2x_3 + 2x_4 = 0 \Rightarrow x_1 + 2x_2 - 4x_4 + 2x_4 = 0 \Rightarrow x_1 = -2x_2 + 2x_4$$

$$2x_3 + 4x_4 = 0 \Rightarrow x_3 = -2x_4$$

- The solution of the above system is of the form:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_2 + 2x_4 \\ x_2 \\ -2x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

- x_2 and x_4 can take any values (free variables).

Computing the Nullspace

- By assigning the value of 1 to a particular free variable and the value of 0 to the rest of the free variables we obtain a so called **special solution**.
- First Special Solution is obtained for $x_2 = 1$, $x_4 = 0$ and is

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

- Second Special Solution is obtained for $x_2 = 0$, $x_4 = 1$ and is

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

- **The null space is the linear combination of the special solutions:**

$$c\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + d\begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

Range and Null Space for linear filtering

- Let's go back to convolution and linear filtering
- We describe it using the Toeplitz matrix $\boldsymbol{A}$: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
- We may want to be able to retrieve $\boldsymbol{x}$ from $\boldsymbol{y}$ (deconvolution)
- The range and the null space of $\boldsymbol{A}$ tells us whether this is possible
- Claim (without proof): If $\boldsymbol{A}$ models a linear convolution then the null space is trivial, otherwise it is normally non-trivial (i.e., for circulant convolution)
- Example: Consider the case of finite difference. The vector $\boldsymbol{x} = (1 \quad \cdots \quad \cdots \quad 1)^T$ is in the null space of $\boldsymbol{A}$ when $\boldsymbol{A}$ is circulant