CT

Complex Numbers: $z\bar{z} \in \mathbb{R}$, $z\bar{z} = a^2 + b^2 \geq 0$, $z\bar{z} = 0 \iff z = 0$ **Standard Inner Product on $\mathbb{C}^n$**: $\forall u, v \in \mathbb{C}^n \quad \langle u, v\rangle = \overline{u}^T v = \sum_{i=1}^{n} \overline{u_i} v_i$

Complex conjugate: $z = a + bi \rightarrow \bar{z} = a - bi$, $z = \bar{z} \iff z \in \mathbb{R}$ **Standard Norm**: $\|u\| = \sqrt{\langle u, u\rangle} = \sqrt{\overline{u}^T u}$

Orthogonal Matrix:
$Q \in \mathbb{R}^{n \times n}$ iff $Q^{-1} = Q^T$

$A = \frac{\sqrt{2}}{2}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$ $B = \begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$

Symmetric Matrices:
A is symmetric if $A^T = A$, and if $A \in \mathbb{R}^{n \times n}$ all eigenvalues λ_i of **A** are real.
All λ_i have algebraic multiplicity = geometric multiplicity. All eigenvectors for distinct eigenvalues are orthogonal.

Vector Norms:
Properties: $\|x\| > 0$ $\|\lambda x\| = |\lambda|\|x\|$ $\|x + y\| \leq \|x\| + \|y\|$

l_p-norm: + friends $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

l_1, l_2, and l_∞ norms are equivalent

Equivalence: for r, s > 0, the l_a and l_b norms are equivalent iff:
$\forall x \in \mathbb{R}^n, r\|x\|_a \leq \|x\|_b \leq s\|x\|_a$ $\forall x \in \mathbb{R}^n$, (1) $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$
(2) $\|x\|_2 \leq \sqrt{n}\|x\|_\infty$
(3) $\|x\|_1 \leq \sqrt{n}\|x\|_2$

Matrix Norms:
Properties: same as above +
$\|AB\| \leq \|A\|\|B\|$ (sub-multiplicative)
Norms: largest of abs sum of cols $\|A\|_1 = \max_j \|a_j\|_1$ largest singular value of **A** $\|A\|_2 = \sigma_1(A)$ largest of abs sum of rows $\|A\|_\infty = \max_i \|a^i\|_1$

Matrix norm || . || on $\mathbb{R}^{m \times n}$ is consistent with the vector norms $||\,.\,||_a$ on $\mathbb{R}^n$ and $||\,.\,||_b$ on $\mathbb{R}^m$ if $\forall$ **$A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$**:
$\|Ax\|_b \leq \|A\|\|x\|_a$ if a = b, || . || compatible with $||\,.\,||_a$

Subordinate Matrix Norm:

$$\forall A \in \mathbb{R}^{m \times n}, \|A\| = \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\}$$
$$\forall A \in \mathbb{R}^{m \times n}, \|A\| = \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\}$$
$$= \max\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n, x \neq 0\}$$
$$= \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| \leq 1\}$$

Frobenius Norm $\|A\|_F \equiv \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}$

For p = 1, 2, ∞, matrix norm $||\,.\,||_p$ is subordinate and compatible with vector norm $||\,.\,||_p$

Orthogonal Projection on a Subspace:

$$\pi_U : \mathbb{R}^m \to \mathbb{R}^m$$
$$v \mapsto \pi_U(v) = U(U^T U)^{-1} U^T v$$
$$\mathrm{im}(A) \perp \ker(A^T)$$

For $A \in \mathbb{R}^{m \times n}$ all vectors $b \in \mathbb{R}^m$, there exists a unique $b_i \in \mathrm{im}(A)$, and a unique $b_k \in \ker(A^T)$ such that $b = b_i + b_k$.

Least Square Method + Linear Regression:
If **Ax = b** has no solution we attempt to minimise $||Ax - b||_2^2$.
Normal Equation: $A^TAx = A^Tb$ gives solution to the least square problem.
Finding $s_0 \in \mathbb{R}$ and $s \in \mathbb{R}^n$ minimising the sum of errors between the model predictions $s_0 + s \cdot a_i$ and observations y_i, can be done by finding the $z = [s_0 \ldots s_n]^T$ minimising $||Az - y||_2^2$, i.e. by solving the normal equation $A^TAz = A^Ty$.

$$\|Ax - b\|_2 \text{ is minimised}$$
$$\iff \|Ax - b_i\|_2 = 0 \iff Ax = b_i$$
$$\iff A^T A x = A^T b$$

Orthogonal transformations preserve the Euclidian length of any vector, and the magnitude of the angle between any two vectors: $\forall u \in \mathbb{R}^n,\ \|Qu\|_2 = \|u\|_2$

$$\forall u, v \in \mathbb{R}^n,\ \widehat{Qu\,Qv} = \widehat{u\,v}$$

det Q = 1 or -1, for all eigvals λ_i, $|\lambda_i| = 1$

Gram-Schmidt Bullshit:

original basis $\vec{v}_1, \vec{v}_2, \ldots \vec{v}_n \longrightarrow$ orthogonal basis $\vec{u}_1, \vec{u}_2, \ldots \vec{u}_n$

1) $\vec{u}_1 = \vec{v}_1$

2) $\vec{u}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1\rangle}{|\vec{u}_1|^2}\vec{u}_1$ $\quad \mathrm{proj}_u(v) \overset{\mathrm{def}}{=} \frac{u \cdot v}{u \cdot u} u$

3) $\vec{u}_3 = \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{u}_1\rangle}{|\vec{u}_1|^2}\vec{u}_1 - \frac{\langle \vec{v}_3, \vec{u}_2\rangle}{|\vec{u}_2|^2}\vec{u}_2$

<v, u> = v . u (see above for complex version)

Spectral Theorem:
For symmetric $A \in \mathbb{R}^{n \times n}$, **A** can be diagonalised as $A = QDQ^T = QDQ^{-1}$ where **Q** is orthogonal and **D** is a diagonal matrix where the diagonal elements are **A**'s eigenvalues.
1. Roots of **det (A – λI)** (eigenvalues $(\lambda_i)_{1 \leq i}$)
2. For each λ_i find the corresponding eigenspace (sub into **A – λI** and solve = 0)
3. Make orthonormal (magnitude 1, and may need **gram-schmidt**)
4. Combine bases to form **Q**
5. Write associated eigenvalues for cols of **Q** in order to form **D**

Positive (Semi-)Definite Matrices:
Definite iff $\forall x \in \mathbb{R}^n - \{0\},\ x^T A x > 0$
- all eigenvalues strictly positive
- all diagonal elements strictly positive
- $\max(A_{ii}, A_{jj}) > |A_{ij}|$

Semi iff $\forall x \in \mathbb{R}^n,\ x^T A x \geq 0$
- all eigenvalues are non-negative
- all diagonal elements non-negative
- $\max(A_{ii}, A_{jj}) \geq |A_{ij}|$

Largest coefficient of **A** on diagonal for *both* definite and semi-definite
Negative equivalents with flipped signs
For matrix **A**, A^TA and AA^T are both symmetric and positive semi-definite

Lower triangular iff $\forall i < j, \quad A_{ij} = 0$
Upper triangular iff $\forall i > j, \quad A_{ij} = 0$

Cholesky Decomposition <3:
$A = LL^T$
A *must* be positive (semi-)definite and symmetric

A is positive semi-definite symmetric matrix $\iff$ there exists a matrix $L \in \mathbb{R}^{n \times n}$ such that: $\begin{cases} L \text{ is lower-triangular} \\ A = LL^T \end{cases}$

A is positive definite symmetric matrix $\iff$ there exists a *unique* matrix $L \in \mathbb{R}^{n \times n}$ such that: $\begin{cases} L \text{ is lower-triangular} \\ A = LL^T \\ \text{the diagonal elements of } L \text{ are positive} \end{cases}$

$L = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$ Solve $A = LL^T$

$$LL^T = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Singular Value Decomposition:
$A = USV^T$
1. Find eigenvalues of AA^T (these form matrix **S** which has the same dimensions as **A** and *descending* sqrt of eigenvalues in the diagonal)
2. Find *orthonormal* set of vectors of A^TA (these are the *columns* of **V** – remember the final product uses V^T!!)
3. Find columns of **U** using formula $u_i = \frac{1}{\sigma_i} A v_i$ for 1 ≤ i ≤ rank(**A**) – remember the V_i come from **V** *not* V^T!! To extend **U** to enough cols pick v_j which are perp to lin comb of existing v_i and G-S.
Mat Dims: **A**: n x m, **U**: n x n, **S**: n x m, V^T: m x m
Properties:
$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$
rank(**A**) = no. +ve singular values in **S**
$\|A\|_2 = \sigma_1$, the largest singular value of A
The positive singular values of **A** are the square roots of the eigenvalues of AA^T or A^TA
Orthonormal Basis for im(A): span of the first rank(**A**) columns of **U**
Orthonormal Basis for ker(A): span of remaining columns of **V** after taking out first rank(**A**) cols

Generalised Eigenvalues:
Given square $A \in \mathbb{R}^{n \times n}$ a non-zero vector $v \in \mathbb{C}^n$ is a generalised eigenvector of rank **m** associated with eigenvalue $\lambda \in \mathbb{C}$ for **A** if:

$$(A - \lambda I)^m v = 0 \text{ and } (A - \lambda I)^{m-1} v \neq 0$$

Thus, any eigenvector associated with **λ** is itself a generalised eigenvector of rank 1.
The image of a vector of the eigenspace
- Associated to **λ** through **A – λI** is **0**.
- of **rank 1** (if there are some) through **A –λI** is in the **eigenspace** associated to **λ**.
- of **rank 2** (if there are some) through **A–λI** is in the **vector space** generated by the **generalised eigenvector** of **rank 1**.
- and so forth...

Generalised EVs Associated with λ:
$A \in \mathbb{R}^{n \times n}$ with eigenvalue $\lambda \in \mathbb{C}$ of algebraic multiplicity **k**, there are k linearly independent generalised eigenvectors $v \in \mathbb{C}^n$ associated with **λ**. It includes the eigenvectors associated with **λ**, as they are generalised eigenvectors.

Number of Generalised Eigenvectors:
$A \in \mathbb{R}^{n \times n}$ has **n** linearly independent generalised eigenvectors. There exist a basis of $\mathbb{C}^n$ of generalised eigenvectors of **A**.

Example:
For matrix **A** defined as: $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
we have $\det(A - \lambda I) = (1 - \lambda)^3$ which gives $\lambda_1 = 1$ and 2 linearly independent EVs $v_1 = (0, 1, -1)^T$, $v_2 = (1, 0, 0)^T$ but since λ_1 has algebraic multiplicity of 3 we find v_3 using $(A - \lambda_1 I)v_3 = v_2$ which gives $v_3 = (0, 0, 1)^T$. We use v_2 here as it is in the row space (a multiple of it is a row of **A**) so v_3 will be linearly independent.

Principal Component Analysis:
$A \in \mathbb{R}^{m \times n}$ = m samples of n dimensional data
$A = USV^T$ - principal axes of **A** = columns of **V**
- principal components of **A** = columns of **US**
Both over 1 ≤ i ≤ rank(**A**)
First principal axis: $w_{(1)} = \arg\max_{\|w\|=1} w^T A^T A w$
Given $A = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$ with $\sigma_1 \geq \ldots \geq \sigma_r > 0$ we see the relation between **A** and the principal components and axes. If $\sigma_1 >> \sigma_2$, the data in **A** can be compressed by projecting in the direction of the principal component: If $\sigma_1 \gg \sigma_2$, then $A \approx \sigma_1 u_1 v_1^T$

This is sometimes used in data compression, PCA, and dimensionality reduction algorithms

Jordan Normal Form:

$$\begin{bmatrix} J_{k_1}(\lambda_1) & & & 0 \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{k_n}(\lambda_n) \end{bmatrix} \quad J_{k_i}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \lambda_i & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_i & 1 \\ 0 & 0 & \ldots & 0 & 0 & \lambda_i \end{bmatrix}$$

Note: The *algebraic* multiplicity of an eigenvalue **λ** is the *sum of the sizes of blocks* with **λ** on the diagonal. The *geometric* multiplicity of **λ** is the *number of blocks* with **λ** on the diagonal.
1. Find eigenvalues of **A** (note the a_i of each **λ**)
2. Find eigenspaces for each (note g_i of each $E_{\lambda i}$)
3. If $g_i < a_i$ find missing $a_i - g_i$ generalised eigenvectors
4. Put eigenvectors in order into matrix (change of basis matrix **B**)
5. $J = B^{-1}AB$

QR Algorithm:
Used to find eigenvalues of matrices, works for most matrices. Consider sequence A_k defined below:
$A_0 = A$
For $k \in \mathbb{N}$ apply the QR decomposition to
A_k: $A_k = Q_{k+1}R_{k+1}$
Stop after sufficient iterations
Properties: (note Q~ is Q with ~ on top denoting orthogonality of Q)
For $k \in \mathbb{N}$, A_k is similar to **A**. $A_k = \tilde{Q}_k^T A \tilde{Q}_k$ so A_k and **A** have the same eigenvalues and **v** is an eigenvector of A_k iff $\tilde{Q}_k v$ is an eigenvector of **A**. The sequence A_k converges to an upper triangular matrix under certain conditions. The eigenvalues of an upper triangular matrix are the diagonal elements.
Symmetric A: All A_k are symmetric. For large enough **k**, the columns of $\tilde{Q}$ are in effect the eigenvectors of **A**.

Fixed Point Equations:
For non-empty set **S** and **f: S → S**, **p ∈ S** is called a fixed point if **f(p) = p**. E.g. for f(x) = x², f(p) = p for p = 0, 1:
$f(p) = p \iff p^2 = p \iff p^2 - p = 0 \iff p(p-1) = 0$
Contraction:
For metric space **(S, d)** and **f : S → S**, **f** is called a contraction of **S** (or a contracting map) if there exists **0 ≤ α < 1** called the contraction constant such that:

$$\forall x, y \in S,\ d(f(x), f(y)) \leq \alpha d(x, y)$$

Fixed Point Theorem:
Let **(S, d)** be a complete metric space and **f** a contraction of **S**. Then **f** has a *unique fixed point*.
Applications: Newton's Method and Initial Value Problem for differential equations.

QR Decomposition (Gram-Shit):
A = QR
$A = [a_1, \ldots, a_n]$, assuming $a_1, \ldots, a_n$ are linearly independent
1. Use Gram-Schmidt to construct an *orthonormal* basis $(e_1, \ldots, e_n)$ s.t.
$\mathrm{span}\,\{e_1, \ldots, e_n\} = \mathrm{span}\,\{a_1, \ldots, a_n\}$
2. $Q = [e_1, \ldots, e_n]$. Note **Q** is semi-orthogonal:

$$Q^T Q = \begin{bmatrix} e_1^T \\ \vdots \\ e_n^T \end{bmatrix} \begin{bmatrix} e_1 & \ldots & e_n \end{bmatrix} = I_{n \times n}$$

3. Choose $R = \begin{pmatrix} (e_1 \cdot a_1) & (e_1 \cdot a_2) & \cdots & (e_1 \cdot a_m) \\ 0 & (e_2 \cdot a_2) & \cdots & (e_2 \cdot a_m) \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & (e_m \cdot a_m) \end{pmatrix}$

Householder Map:
For hyper-plane **P** going through the origin with unit normal $u \in \mathbb{R}^m$, i.e.,
P = {x ∈ R m : u · x = 0} the Householder matrix defined by $H_u = I - 2uu^T$ induces reflection wrt **P**.
Properties:
- H_u is involutory: $H_u = H_u^{-1}$
- H_u is orthogonal: $H_u^T = H_u^{-1}$
- H_u preserves the Euclidian length of vectors: $||H_u(x)|| = ||x||$
- H_u preserves angles between vectors
- All rotations and reflections are orthogonal operations
- Orthogonal projection **Q** on the hyperplane **P** is given by: $Q = I - uu^T$ with $Q^2 = Q$ and $Q = Q^T$

Convergence of QR in A = LU Case:
Non-singular matrix **A** can be factorised as **A = LU** where **L** is lower triangular and **U** upper, iff **A** can be reduced to REF without swapping any rows. Transform **A** into REF to get **U**, then we know $L^{-1}A = U$ so can find **L**.
Uniqueness: **A = LU** is only unique iff **A** is non singular and the diagonal elements of **L** are all 1.
$A^n = Q_1 \ldots Q_n R_1 \ldots R_n$
Convergence: Let $A \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix with distinct eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$ with eigendecomposition $A = Q\Lambda Q^T$. Suppose $Q^T = LU$ with unit lower triangular **L** and the diagonal elements of **U** are positive. Then $A_k \to \Lambda$

Convergence:
Convergence of real numbers:
$\lim_{n \to \infty} a_n = l$ iff $\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ such that $\forall n > N,\ |a_n - l| < \epsilon$
1. Find limit l
2. Take ε > 0
3. Put $|a_n - l| < \varepsilon$, find expression for n > … and set N = roof of what n is >
Cauchy Sequence: $\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ such that $\forall n, m > N,\ |a_n - a_m| < \epsilon$ (terms get gradually closer). a_n is only convergent if it is Cauchy.
Metric Spaces:
A tuple **(S, d)** where **S** is a non-empty set and **d** is a metric over **S** (**d : S x S → R**).
Prove the below properties hold to show we have a metric space:
1. $\forall x, y \in S,\ d(x, y) \geq 0$
2. $\forall x, y \in S,\ d(x, y) = 0 \iff x = y$
3. $\forall x, y \in S,\ d(x, y) = d(y, x)$
4. $\forall x, y, z \in S,\ d(x, y) \leq d(x, z) + d(z, y)$

Convergence in a Metric Space:
Space = **(S, d)**, sequence = a_n, limit = **l ∈ S**
a_n converges to **l** iff:
$\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ such that $\forall n > N,\ d(a_n, l) < \epsilon$
If a_n converges it's limit is unique.

Cauchy Seq. (Metric Spaces): $\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ such that $\forall n, m > N,\ d(a_n, a_m) < \epsilon$
For **(S, d)** and a_n a sequence in **S**, a_n is only convergent if it is Cauchy.
Complete Spaces:
Metric space **(S, d)** is a complete space iff every Cauchy sequence in **S** is also converging in **S**. For any **k > 0**, $\mathbb{R}_k$ equipped with any of the three metrics induced by l_1, l_2 or l_∞ norms is complete.

Condition Number:
Measure of sensibility of a problem to small fluctuations in input.
Stability of the System: how the system responds to noise in input
Sensibility of Solution: how small changes in parameters affect the solution of a parametric equation
Let **P** = problem of interest, **d** = input, **ε** = perturbation in input, **s(d)** = desired output, **s(d + ε)** = perturbed output.

$$\kappa(P) = \max_{\epsilon} \frac{\|s(d) - s(d+\epsilon)\|}{\|\epsilon\|}$$

Relative Condition Number:

$$\kappa(P) = \max_{\epsilon} \frac{\|s(d) - s(d+\epsilon)\|}{\|\epsilon\|} \frac{\|d\|}{\|s(d)\|}$$

Value depends on the norms being used. Can also be defined in terms of relative difference.
Unstable system/Ill-conditioned problem: large condition number (cannot always assume ill-conditioned for matrices though!!)
Stable system/Well-conditioned problem: small condition number
Condition Number for Square Non-Singular Matrices:

$$\boldsymbol{\kappa(A) = \|A^{-1}\|\|A\|}$$

Gives a bound on the relative change on A^{-1} given by perturbation on **A**. $\kappa(A) \geq 1$ and $\kappa(\gamma A) = \kappa(A)$
Pseudo-Inverse and Condition Number:
Not all matrices can be inverted hence **k(A)** cannot be calculated. We can instead use the pseudo-inverse and define a generalisation of the above formula for **k(A)**:

$$\kappa(A) = \|A^{\dagger}\|\|A\| \quad A^{\dagger} = (A^TA)^{-1}A^T$$

A^TA is square, then $Ax = b \rightarrow A^TAx = A^Tb \rightarrow x = (A^TA)^{-1}A^Tb$

Iterative Solutions of Linear Equations:
Jacobi: each stage uses the previous stage's results
Gauss-Seidel: each stage uses most recent values
G-S Convergence: For **Ax = b**, converges if

$$|a_{i,i}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{i,j}|$$

Simultaneous equations
$u(x_1, x_2)$ and $v(x_1, x_2)$ converge if:

$$\left|\frac{\partial u}{\partial x_1}\right| + \left|\frac{\partial u}{\partial x_2}\right| < 1 \quad \text{and} \quad \left|\frac{\partial v}{\partial x_1}\right| + \left|\frac{\partial v}{\partial x_2}\right| < 1$$

Prove G-S convergence using 2x2 general **A** and **b**
Splitting – General Method:
A = G + R, then solve $x_{k+1} = Mx_k + c$. For consistent norm **||.||** on $R^{n \times n}$ if **||M|| < 1** the sequence converges for any starting point x_0.
Rate of convergence $r \propto -\log_{10}||M||$
Note about non-diagonal matrices:
For a matrix like $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with all diagonal elements 0, we cannot split it so we use a change of basis $C = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ ($C = C^{-1}$):

$$Ax = b \Leftrightarrow \quad C^{-1}Ax = C^{-1}b \Leftrightarrow \quad (C^{-1}AC)C^{-1}x = C^{-1}b$$

By denoting $C^{-1}AC = B$, $C^{-1}x = y$, and $C^{-1}b = c$ we solve **By = c** and can retrieve **x** through **x = Cy**.
Jacobi Method:
A = D + R where **R = L + U** (**D** = diags, **L** = lower tri, **U** = upper tri)
$M = -D^{-1}R$ and $c = D^{-1}b$
Gauss-Seidel Method:
A = (D + L) + U, $M = -(D + L)^{-1}U$, $c = (D + L)^{-1}b$
Condition Number and Convergence:
For **Ax = b** if the condition number of **A** is large J and G-S converge more slowly/not at all
If **A** is weakly row diagonally dominant and irreducible, J and G-S both converge (G-S faster)

Diagonally Dominant Matrix:
$A \in R^{n \times n}$ is strictly row diagonally dominant if:

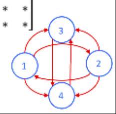
$$\forall i, |a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$$

A is non-singular.
Let $A \in R^{n \times n}$ and $\rho(A) = \max_{\lambda \in Sp(A)} |\lambda|$ be the spectral radius of **A.**
For $\epsilon > 0$, there exists an induced norm such that $||A|| < \rho(A) + \epsilon$.

Irreducible Matrix:
If **A** (an n x n matrix) cannot take the form $P^TAP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ through symmetric permutation of its rows and columns, where A_{11} etc are square block matrices and **P** is a permutation matrix.
Permutation Matrix:
Square matrix with all elements 0 except exactly one 1 in each row/column. $P^TP = PP^T = I$.
Checking Irreducibility with Graph:

$$A = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix}$$

If a_{ij} where , **i ≠ j**, draw an arrow from point **i** to **j**.
Irreducible iff from any point you can go to any other point by following arrows.
Given graph is for a **4x4 reducible matrix**.



Gradient Based Optimisation:
Quadratic Form: $Q(x_1, x_2) = x_1^2 + 2x_1x_2 - 3x_2x_1 + 5x_2^2$

$$Q(x_1,x_2) = [x_1 \quad x_2]^T \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad Q(x_1,x_2) = [x_1 \quad x_2]^T \begin{bmatrix} 1 & -0.5 \\ -0.5 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

For symmetric **A**: $f(x_1, x_2) = x^TAx - 2x^Tb$
Gradient/Steepest Descent: Finds local minimum
<u>Gradient</u>: $\vec{x}^{(k+1)} = \vec{x}^{(k)} - \alpha\nabla f(\vec{x}^{(k)})$ where α is a fixed step size
Find α: $x^{(1)} = x^{(0)} + hg^{(0)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + h\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 + 4h \\ 1 + 4h \end{bmatrix}$ ←Exact line search
Alt: choose const, diminishing step size (1/k)
Sub x_1 into f(x) and solve d/dh = 0 to get optimal h (= α)
<u>Steepest</u>: $\vec{x}^{(k+1)} = \vec{x}^{(k)} - \alpha_k\nabla f(\vec{x}^{(k)})$, $\alpha_k = \arg\min f\left(\vec{x}^{(k)} - \alpha\nabla f(\vec{x}^{(k)})\right)$
For quadratic function $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x} - b^T\boldsymbol{x}$, $\quad \boldsymbol{\alpha_k = \frac{(g^{(k)})^T(g^{(k)})}{((g^{(k)})^T)Hg^{(k)}}}$
where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x})$ and $H = Q = \nabla^2 f(\boldsymbol{x})$
Remember works in orthogonal steps, so $(x^2 - x^1) . (x^3 - x^2) = 0$
Gradient Ascent: Finds local maximum

Conditioning of a Problem:
T-Digit Arithmetic:
Rule of thumb: for condition number **k(A)** you lose about $c = \log_{10} k(A)$ significant figures in accuracy
Error Bounds and Iterative Refinement:
When solving **Ax = b**, with approximate solution **x~**, using the residual vector **||r|| = ||b = Ax~||** or other geometric measures may not always be good ways to measure how good an estimate **x~** is. The norms of **A** and A^{-1} may provide useful information on error bounds through the following theorem:
Suppose that **x~** is an approximation to the solution of **Ax = b**, and **A** is a non-singular matrix and **r** is the residual vector of **x~**. Then for any natural norm $\|x - \tilde{x}\| \leq \|r\|\|A^{-1}\|$ and if x ≠ 0 and b ≠ 0,

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|} = \kappa(A)\frac{\|r\|}{\|b\|}$$

E.g. for **A** we have $||A||_\infty = 3.0001$, $||A^{-1}||_\infty = 20000$, and $k(A) = ||A||_\infty ||A^{-1}||_\infty = \mathbf{60002}$ – the size of the condition number should keep us away from making hasty decisions on accuracy.

Iterative Techniques for Eigenvalues and Eigenvectors:
Power Method/Power Iteration:
Take initial vector $x_0 = [1, 1, 1, 1]^T$. Converges to the eigenvalue with largest modulus.

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|} \text{ and } x_0 \in \mathbb{R}^n \backslash \{0\}$$

$$x_k \xrightarrow[k\to\infty]{} v \quad \text{and} \quad \|Ax_k\| \xrightarrow[k\to\infty]{} |\lambda|$$

Limitations: If not all eigenvalues have distinct modulus, will return linear combination of the corresponding EVs. Convergence slow if dominant eigenvector not very dominant.
Inverse Power Iteration: Do the above on A^{-1} to get $1/\lambda$ where λ is the EV with smallest modulus of **A**.
Shifts: A – sI
$\lambda \in R$ is an eigenvalue of matrix **A** iff $(\lambda - s)$ is an eigenvalue of matrix **A – sI**. If $\{\lambda, v\}$ is an eigenpair of **A** and $s \neq \lambda$ then $\{1/(\lambda - s), v\}$ is an eigenpair of $(A - sI)^{-1}$.
Allows us to focus on other eigenvalues.
Method: choose **s** such that $\mu_1 = 1/(\lambda_j - s)$ is the dominant eigenvalue of $(A - sI)^{-1}$. The eigenvalue of interest for **A** is given by : $\lambda_j = 1/\mu_1 + s$. If the eigenvector oscillates, $1/\mu_1$ should be negative.
Rayleigh Quotient:
While using an iterative technique, can use Rayleigh Quotient to monitor convergence to eigenvalue directly, not it's modulus:

$$R(A,x) = \frac{x^TAx}{x^Tx}$$

Deflation:
Find second dominant eigenvalue by deflating **A** to $B \in R^{(n-1) \times (n-1)}$ which has all eigenvalues of **A** except dominant one.
Steps: find $u = x_1 + \|x_1\|_2 e_1$, do $H = I - \frac{2uu^T}{u^Tu}$, then $HAH^{-1} = \begin{bmatrix} \lambda_1 & b^T \\ 0 & B \end{bmatrix}$ (remember $H^{-1} = H$) Then can use **B** to get λ_2

Linear Programming:
Graphical Method:
Draw lines, shade exclusion zones, rearrange objective function for y, move line until maximised, sub for P.
Simplex Method: (stop when no more –ve in **z** row)
1. Most negative value in **z** row (highlight col)
2. Ratios (sol / col), highlight lowest ratio's row
3. Replace row var (basic) with highlighted col var (non-basic)
4. Make pivot = 1 (divide highlighted row by cross highlighted value
5. Gaussian elimination to clear rest of highlighted col

Conjugate Gradient Method:
Ax = b, for $A \in R^{n \times n}$ converge in **n** steps
Larger condition number = slower convergence
Check convergence by size of norm of residual

Iteration 1:
Use $r_0 = b - Ax_0$ as d_1
$\alpha_1 = (r_0^Tr_0) / (d_1^TAd_1)$
$x_1 = x_0 + \alpha_1 d_1$
$r_1 = r_0 - \alpha_1 Ad_1$

Iteration k > 1:
$\beta_k = -(r_{k-1}^Tr_{k-1}) / (r_{k-2}^Tr_{k-2})$
$d_k = r_{k-1} - \beta_k d_{k-1}$
$\alpha_k = (r_{k-1}^Tr_{k-1}) / (d_k^TAd_k)$
$x_k = x_{k-1} + \alpha_k d_k$
$r_k = r_{k-1} - \alpha_k Ad_k$

Common Stopping Criteria:
Gradient close enough to 0 $\|\nabla f\| < \epsilon$
Improvements in function value are saturating

$$|f(x_{k+1}) - f(x_k)| < \epsilon$$

Movement between iterates small enough

$$\|x_{k+1} - x_k\| < \epsilon$$

Relative measure (no dependence on scale of **f**)

$$\frac{|f(x_{k+1}) - f(x_k)|}{\max(1, |f(x_k)|)} < \epsilon$$

Another relative measure $\frac{\|x_{k+1} - x_k\|}{\max(1, \|x_k\|)} < \epsilon$

Algorithm for Solving System of Linear Equations:

```
r = b − Ax
repeat in the loop
    α = rᵀr / rᵀAr
    x = x + αr
    exit if ||r|| is sufficiently small
    r = r − αAr
end repeat loop
return x as the result
```

Note that **x = x + αr**
r (residual) = b – Ax
= b – A(x + αr) = r - αAr
Slow convergence hence not widely used for linear equs – used for non-linear equs tho

Iterative Refinement:
Solving **Ax = b** numerically you get **x~**, with the residual vector defined as **r = b – Ax~** being a reliable indicator of accuracy iff **A** is well-conditioned. Iterative refinement aims to reduce round-off errors. Suppose we are solving:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}$$

Approx sol: $[\widetilde{x_1} \ \widetilde{x_2} \ \widetilde{x_3}]$
Correction factors: $[\Delta x_1, \Delta x_2, \Delta x_3]$
Where: $\begin{cases} x_1 = \widetilde{x_1} + \Delta x_1 \\ x_2 = \widetilde{x_2} + \Delta x_2 \\ x_3 = \widetilde{x_3} + \Delta x_3 \end{cases}$

We sub the formulae of x_1, x_2, and x_3 into the above, then subtract the subbing in of **x~** as the solution to obtain:

$$\begin{cases} a_{11}\Delta x_1 + a_{12}\Delta x_2 + a_{13}\Delta x_3 = b_1 - \widetilde{b_1} = r_1 \\ a_{21}\Delta x_1 + a_{22}\Delta x_2 + a_{23}\Delta x_3 = b_2 - \widetilde{b_2} = r_2 \\ a_{31}\Delta x_1 + a_{32}\Delta x_2 + a_{33}\Delta x_3 = b_3 - \widetilde{b_3} = r_3 \end{cases}$$

We solve to obtain new correction factors to improve the solution of x_i.
With t-digit arithmetic and Gaussian Elimination, one can expect $\|r\| \approx 10^{-t}\|A\|\|\tilde{x}\|$ and one can show that the approximation to the condition number is: $\kappa(A) \approx \frac{\|x - \tilde{x}\|}{\|\tilde{x}\|} 10^t$

Functions of Several Variables:
Clairaut's Theorem:
$f_{xy}(a, b) = f_{yx}(a, b)$

$$(f_x)_x = f_{xx} = f_{11} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2}$$

$$(f_x)_y = f_{xy} = f_{12} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 z}{\partial y \partial x}$$

$$(f_y)_x = f_{yx} = f_{21} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 z}{\partial x \partial y}$$

$$(f_y)_y = f_{yy} = f_{22} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2}$$

Directional Derivatives:
If f is a differentiable function of x and y then **f** has a directional derivative in the direction of unit vector **u = <a, b>**: $D_uf(x, y) = f_x(x, y)a + f_y(x, y)b$. For unit vector at angle **θ** use **<a, b> = <cos θ, sin θ>** (maximised at **θ = 0**). D_u represents the rate of change of **z** in the direction of **u**. Can also be

$$D_{\mathbf{u}}f(x,y) = \langle f_x(x,y), f_y(x,y)\rangle \cdot \mathbf{u} \text{ , } \nabla f(x,y) = \langle f_x(x,y), f_y(x,y)\rangle = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$

Directional derivative = scalar proj of gradient vec onto **u** (make **u** a unit vector first).

Tangent Plane and Normal Line to Level Surface:
If $\nabla F(x_0, y_0, z_0) \neq 0$, it is natural to define the tangent plane to the level surface $F(x, y, z) = k$ at $P(x_0, y_0, z_0)$ as the plane that passes through **P** and has normal vector $\nabla F(x_0, y_0, z_0)$. $\nabla F(x_0, y_0, z_0)$ gives the direction of fastest increase of **f**.

Maxima and Minima:
Local max: at **(a, b)** if **f(x, y) ≤ f(a, b)** for all points **(x, y)** in some disk with centre **(a, b)**. Similar for local min and global max/min if inequalities hold for all **(x, y)**. $f_x(a, b) = 0$ and $f_y(a, b) = 0$ ($\nabla f(a, b) =$ **0**). Critical point if prev condition or one does not exist.
Second Derivatives Test:

$$\boldsymbol{D = D(a,b) = f_{xx}(a,b)f_{yy}(a,b) - [f_{xy}(a,b)]^2}$$

a. If $D > 0$ and $f_{xx}(a,b) > 0$, then $f(a,b)$ is a local minimum
b. If $D > 0$ and $f_{xx}(a,b) < 0$, then $f(a,b)$ is a local maximum
c. If $D < 0$, then $f(a,b)$ is not an extrema but a saddle point
d. If $D = 0$, then the test is inconclusive and $f(a,b)$ could be a local maxima, minima or a saddle point.

Hessian Matrix: $\boldsymbol{D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2}$ where $\boldsymbol{f_{xy} = f_{yx}}$
2nd derivative test generalizes to test based on eigenvalues of **D** (+ve = min, -ve = max, mix = saddle, singular **D** = inconclusive).

Find the maximum of

$$z = 5x_1 + 4x_2 + 6x_3$$

subject to constraints

$$\begin{aligned} 4x_1 + x_2 + x_3 &\leq 19 \\ 3x_1 + 4x_2 + 6x_3 &\leq 30 \\ 2x_1 + 4x_2 + x_3 &\leq 25 \\ x_1 + x_2 + 2x_3 &\leq 15 \\ x_1, \ x_2, \ x_3 &\geq 0 \end{aligned}$$

Maximise $z = 5x_1 + 4x_2 + 6x_4$

$$\begin{aligned} 4x_1 + x_2 + x_3 + x_4 &= 19 \\ 3x_1 + 4x_2 + 6x_3 + x_5 &= 30 \\ 2x_1 + 4x_2 + x_3 + x_6 &= 25 \\ x_1 + x_2 + 2x_3 + x_7 &= 15 \\ x_1, x_2, x_3, x_4, x_5, x_6, x_7 &\geq 0 \end{aligned}$$

1	Non-basic variables			Basic variables				
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Solution
z	−5	−4	−6	0	0	0	0	0
x_4	4	1	1	1	0	0	0	19
x_5	3	4	6	0	1	0	0	30
x_6	2	4	1	0	0	1	0	25
x_7	1	1	2	0	0	0	1	15

2	Non-basic variables			Basic variables					
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Solution	Ratios
z	−5	−4	−6	0	0	0	0	0	
x_4	4	1	1	1	0	0	0	19	19/1 = 19
x_5	3	4	6	0	1	0	0	30	30/6 = 5
x_6	2	4	1	0	0	1	0	25	25/1 = 25
x_7	1	1	2	0	0	0	1	15	15/2 = 7.5

3	Non-basic variables			Basic variables					
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Solution	Ratios
z	−5	−4	−6	0	0	0	0	0	
x_4	4	1	1	1	0	0	0	19	19/1 = 19
x_3	3/6	4/6	6/6	0	1/6	0	0	30/6	30/6 = 5
x_6	2	4	1	0	0	1	0	25	25/1 = 25
x_7	1	1	2	0	0	0	1	15	15/2 = 7.5

4	Non-basic variables			Basic variables					
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Solution	Ratios
z	−2	0	0	0	1	0	0	30	
x_4	7/2	1/3	0	1	−1/6	0	0	14	14/(7/2)=4
x_3	1/2	2/3	1	0	1/6	0	0	5	5/(1/2)=10
x_6	3/2	10/3	0	0	−1/6	1	0	20	20/(3/2)=13.3
x_7	0	−1/3	0	0	−1/3	0	1	5	-

5	Non-basic variables			Basic variables				
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Solution
z	0	4/21	0	4/7	19/21	0	0	38
x_1	1	2/21	0	2/7	−1/21	0	0	4
x_3	0	13/21	1	−1/7	4/21	0	0	3
x_6	0	67/21	0	−3/7	−2/21	1	0	14
x_7	0	−1/3	0	0	−1/3	0	1	5

Solution:
$x_1 = 4$, $x_2 = 0$, $x_3 = 3$, $z = 38$
When x_n not in z col, $x_n = 0$.

Dual LP Problem:
Every minimisation/maximisation has a dual problem which aims to max/min based on the same constraints. When both are optimised they are equal. Sometimes the dual is easier to solve than the primal.

Primal problem
Minimise: $Z = 12x_1 + 16x_2$
Subject to the constraints:
$x_1 + 2x_2 \geq 40$
$x_1 + x_2 \geq 30$
$x_1 \geq 0; \ x_2 \geq 0$

Dual problem
Maximise: $Z = 40y_1 + 30y_2$
Subject to the constraints:
$y_1 + y_2 \leq 12$
$2y_1 + y_2 \leq 16$
$y_1 \geq 0; \ y_2 \geq 0$

Original matrix (minimisation)		
1	2	40
1	1	30
12	16	0

Transpose matrix (maximisation)		
1	1	12
2	1	16
40	30	0

Then solve maximisation problem and get the same result :)