Primary Decomposition Theorem :-

Let $T \in L(V)$ and $m_T(x) = (x-\lambda_1)^{s_1} \cdots (x-\lambda_k)^{s_k}$, then

$$V = \ker(T-\lambda_1 I)^{s_1} \oplus \cdots + \ker(T-\lambda_k I)^{s_k}$$

and each $\ker(T-\lambda_i I)^{s_i}$ is a T-invariant subspace of V.

Result :- Eigen vectors corresponding to distinct eigen values of a Linear operator T are linearly independent.

Diagonalizability :-

A linear operator $T \in L(V)$ is said to be diagonalizable if $\exists$ an invertible matrix P such that $P^{-1}[T]P$ is a diagonal matrix D $\left[P^{-1}[T]P = D \quad \text{or} \; [T] = PDP^{-1} \right]$

In this case, the diagonal elements of D are the eigen values of T and the columns of P are the eigen vectors of T.

If all the eigen values of T are distinct then the corresponding eigen vectors will be linearly independent, so, in this case D will be invertible, and hence T will be diagonalizable.

Geometric Multiplicity :- dimension of eigen-space $\ker(T-\lambda I)$

Algebraic Multiplicity :- Multiplicity of λ as a root of the characteristic polynomial of T.

The $GM \leq AM$. for any eigenvalue of λ.

Q. When T is diagonalizable?

The following are equivalent —

(i) T is diagonalizable

(ii) $GM(\lambda) = AM(\lambda)$ for every eigen value of λ

(iii) $V = \ker(T-\lambda_1 I) \oplus \ker(T-\lambda_2 I) + \cdots + \ker(T-\lambda_k I)$

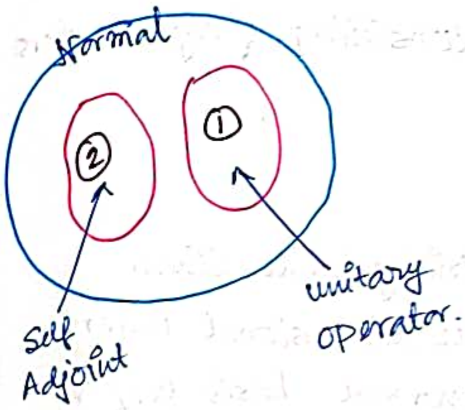
(iv) $m_T(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$

Jordan Canonical form :-

If T is not diagonalizable, we can still put T in a "almost diagonal" form called Jordan Canonical form.

① Unitary operator :- $TT^* = T^*T = I$ $\quad [AA^* = A^*A = I_{n\times n}]$

Properties ① TFAE : (i) T is unitary
(ii) T preserves inner products
(iii) T preserves norm
(iv) T maps orthonormal bases to orthonormal bases.



② E-values of a unitary operator have absolute value 1.

③ for $A \in F^{n\times n}$

TFAE: ① A is unitary
② The columns of A form an orthonormal bases of F^n.

④ The change of Basis matrix $[T]_{B_2 B_1}$ for orthonormal Bases B_1 & B_2 is a unitary matrix.

Normal Operator :- definition $TT^* = T^*T$.

Properties :- ① T is normal $\Leftrightarrow \|Tv\| = \|T^*v\|.\ \forall v \in V$.
② $Tv = \lambda v \Rightarrow T^*v = \bar{\lambda} v$.
③ Eigen vectors corresponding to distinct eigen values are orthogonal
④ If $T^k v = 0$ for some $k \geq 1$ then $Tv = 0$
[i.e. if $v \in \ker T^k$ then $v \in \ker T$; thus $\ker T^k = \ker T$ for any $k \geq 1$]

② Self Adjoint operator :- $T = T^*$ [thus $T^*T = TT^*$ so, T is normal]

① For every $v \in V$, $\langle Tv, v\rangle$ is real.
② $\langle Tv, v\rangle = 0 \quad \forall v \in V \Rightarrow T = 0$
③ If $T^k v = 0$ for some $k \geq 1$, then $Tv = 0$
(i.e $\ker T^k = \ker T$)
④ All e-values of T are Real. ⑤ eigen vectors corresponding to distinct eigen values are orthogonal.

Spectral Theorem :- [Spectrum of T := set of eigen values of T]

Let T be a triangulable linear operator on a finite dimensional inner product space V. Then T is normal $\Leftrightarrow$ V has an orthonormal basis consisting of eigenvectors of T.

Schur's Theorem :- Let T be a triangulable linear operator on a finite dimensional inner product space V. Then $\exists$ an orthonormal basis B of V such that the matrix of T w.r.t B is upper triangular.

Corollary :- If A is a unitary upper triangular matrix, then $A^{-1} = A^*$

hence A will be diagonal matrix with diagonal entries having absolute value 1.

Singular Value Decomposition :-

Given $A \in \mathbb{C}^{m \times n}$, a singular value decomposition of A is a factorization $A = U \Sigma V^*$ where U & V are unitary & Σ is diagonal.

A positive definite matrix is a symmetric matrix whose eigen values are positive.

SVD can be looked at from three mutually compatible points of view —

① We can see it as a method for transforming correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items

② SVD is a method for identifying and ordering the dimensions along which data points exhibit the most variation.

③ once we have identified where the most variation is, it's possible to find the best approximation of the original data points using fewer dimensions.

— SVD — Method for Data Reduction.

$$\boxed{A_i v_i = \sigma_i u_i}$$

2 Subclasses of self adjoint operators:-

Positive semi definite

$\langle Tv, v \rangle \geq 0 \quad \forall v \in V$

Equivalent conditions.
- all eigen values are non negative
- $\sqrt{T}$ exists.
- $\exists\ S \in L(V)$ such that $T = S^* S$.

Positive definite

$\langle Tv, v \rangle > 0 \quad \forall v \in V$

- all eigen values are positive
- positive definite $\sqrt{T}$ exists
- $\exists$ invertible $S \in L(V)$ s.t $T = S^* S$.

The essential notions of size and distance in a vector space are captured by norms. Measures with which we measure approximations and convergence throughout the LAA.

Matrix Norms :- Similar to vector norm we have matrix norms

- vector Norm : a function $\|\cdot\| : V \rightarrow \mathbb{R}$ satisfying
  - $\|x\| \geq 0$ & $\|x\| = 0 \Leftrightarrow x = 0$ [norm is positive definite]
  - $\|x+y\| \leq \|x\| + \|y\|$ [triangle inequality]
  - $\|\alpha . x\| = |\alpha| . \|x\|$ [homogeneity property]

- An important class of vector norms is the class of p-norms, defined for $p \geq 1$:

$$\|x_p\| = \left[\sum |x_i|^p\right]^{1/p}$$

euclidian norm is a special case of p norm when $p = 2$.

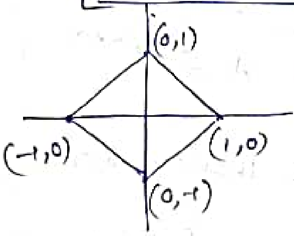
closed unit discs in some p norms | the set of vectors whose norms are equal to 1 |

$p = 1$ $\quad \|x\|_1 = \sum |x_i|$

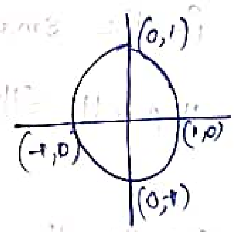
$|x_1| + |x_2|$

say $n = 2$ $\quad x = (x_1, x_2)$

max column sum



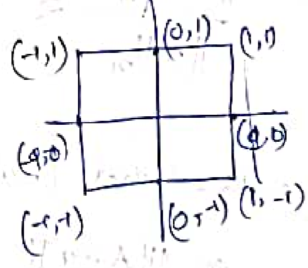
$p = 2$ $\quad \|x\|_2 = \left[\sum |x_i|^2\right]^{1/2} = (x_1^2 + x_2^2)^{\frac{1}{2}}$



$p = \infty$ $\quad \|x\|_\infty = \max_i \{|x_i|\}$ $\quad p = \infty$

max. row sum

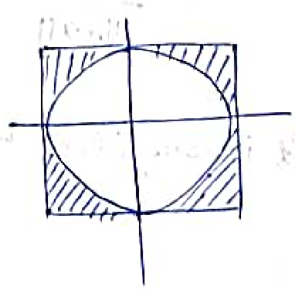
$\max\{|x_1|, |x_2|\}$



If we go on increasing p, we will notice that the circle will converge to the square, that is the inspiration behind the definition because infinity is not a number so we cannot give it in terms of $p = \infty$. that is why we are taking max of $|x_1|$ and $|x_2|$.

$2 \leq p \leq \infty$

the intermediate between 2 and ∞ norm.



Another class : weighted p- norms

$\|x\|_w := \|Wx\|$ for any norm $\|\cdot\|$ & any non- singular matrix W.

Matrix Norms :-

A matrix norm is a function $\|\cdot\| : \mathbb{C}^{m\times n} \rightarrow \mathbb{R}$ satisfying

- $\|A\| \geq 0 \quad \forall A$ & $\|A\| = 0 \iff A = 0$
- $\|A+B\| \leq \|A\| + \|B\|$
- $\|\alpha A\| = |\alpha| . \|A\|$.
- $\|AB\| \leq \|A\| . \|B\|$ [whenever the product of the 2 matrices are defined.]. $\boxed{m=n.}$

Important examples :-

① Induced matrix norms :-

Suppose $A \in \mathbb{C}^{m\times n}$, consider $A : \mathbb{C}^n \rightarrow \mathbb{C}^m$

$\|\cdot\|_n \quad \|\cdot\|_m$

then the induced norm $\|A\|_{(m,n)}$ is the smallest scalar C such that

norm on Cm or Cn

$\|Ax\|_m \leq \|x\|_n . C$

i.e $\dfrac{\|Ax\|_m}{\|x\|_n} \leq C$ [C is the max. factor or largest stretch that A can apply 'stretch' x.].

we know that $\|\alpha x\| = |\alpha| . \|x\|$.

so, $\dfrac{\|A\alpha x\|}{\|\alpha x\|} = \dfrac{|\alpha| \, \|Ax\|}{|\alpha| \, \|x\|} = \dfrac{\|Ax\|}{\|x\|}$

it is sufficient to consider vectors x of norm 1.

$$\|A\|_{(m,n)} = \sup_{\substack{x \neq 0 \\ x \in \mathbb{C}^n}} \frac{\|Ax\|_m}{\|x\|_n}$$

$$= \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_n = 1}} \|Ax\|_m$$

1. Let A be an $m \times n$ matrix

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}$$

$A = [a_0 | a_1 | \cdots | a_{n-1}]$ or $A = [a_1 | a_2 | \cdots | a_n]$

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

$$= \sqrt{\sum_{j=1}^{n} \|a_j\|_2^2}$$

$$= \left[\left\| \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{matrix} \right\|_2^2 \right]^{1/2}$$

it equals the vector-2 norm of the vector that is created by stacking the columns of A on top of each other.

Theorem :- Let A be a $m \times n$ matrix

$$A = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$$

Let a^j denote the jth column of A and

a_i denote the ith row of A.

① $$\|A\|_1 = \max_{1 \le j \le n} \|a^j\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}| \quad \text{(max. column sum)}$$

② $$\|A\|_\infty = \max_{1 \le i \le m} \|a_i\|_1 = \max_i \sum_{j=1}^{n} |a_{ij}| \quad \text{(max. row sum)}$$

③ $$\|A\|_2 = \sqrt{\rho(A^*A)} = \text{largest singular value of } A.$$

The Four Fundamental Subspaces :-

The column space $C(A)$ contains all combinations of the columns of A

The row space $C(A^T)$ contains all combinations of the columns of A^T.

The Null Space $N(A)$ contains all solutions x to $Ax = 0$

The left Null space $N(A^T)$ contains all solutions y to $A^T y = 0$.

$[A]_{m \times n}$

dimension of row space $= r = $ rank

dimension of column space $= r =$ rank

dimension of null space $= (n-r)$

dimension of null space of $A^T = (m-r)$

Floating point representation :-

Floating point = fraction $\times$ base$^{\text{Power}}$

↑ precision accuracy (fraction); ↑ range (Power)

Floating point numbers with base $= 2$

$(-1)^S \times F \times 2^E$.

↑ F: fraction (fixed point number)

Single precision number :-

1	8	23.
S	E	F

total 32 bits.

double precision number :-

1	11	20 + 32
S	E	F

total 64 bits

$5 \times 2^{+4} \Leftrightarrow 1.25 \times 2^{6}$

$f(x) = x^2 - 2x + 1 \Rightarrow \emptyset -$

Positive Definite System :-

A matrix $A \in \mathbb{R}^{n\times n}$ is positive definite if $x^T A x > 0 \quad \forall$ nonzero $x \in \mathbb{R}^n$.

+ve semidefinite if ~~$x^T A$~~ $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$

indefinite if we can find $x, y \in \mathbb{R}^n$ such that

$$(x^T A x).(y^T A y) < 0.$$

Symmetric +ve definite systems constitute one of the most important classes of special $Ax = b$ problem.

Let A is a symmetric matrix

$$A = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$

$x = [0, 1]^T \Rightarrow x^T A x = \alpha > 0$

$y = [1, 0]^T \Rightarrow y^T A y = \gamma > 0$

$x = [1, 1]^T \Rightarrow x^T a x = \alpha + 2\beta + \gamma > 0$

$y = [1, -1]^T \Rightarrow y^T a y = \alpha - 2\beta + \gamma > 0$

from these $\Leftrightarrow \quad \alpha + 2|\beta| + \gamma \geq 0$,

or, $\dfrac{\alpha+\gamma}{2} \geq |\beta|$

The Largest entry in A is on the diagonal and that is +ve.

A symmetric positive definite matrix has a diagonal that is sufficiently "weighty" to ~~pre~~ preclude the need for pivoting.

A special Cholesky factorization is there

Positive definiteness :-

$A \in \mathbb{R}^{n\times n}$ is positive definite. It is obvious that a positive definite matrix is non singular for otherwise we could find x, $x^T A x > 0$ such that $x \neq 0$.

Theorem :- If $A \in \mathbb{R}^{n\times n}$ is positive definite and $X \in \mathbb{R}^{n\times k}$ has rank k. then $B = X^T A X \in \mathbb{R}^{k\times k}$ is also +ve definite

If $z \in \mathbb{R}^k$ satisfies. $z^T B z \leq 0$

$$z^T X^T A X z \leq 0.$$

$$\Rightarrow (Xz)^T A (Xz) \leq 0.$$

then A is positive definite.

So, $Xz = 0$

$\Rightarrow$ X has full column rank.

So, $z = 0$.

then B is also +ve definite.

Corollary :- If A is positive definite, then all its principal submatrices are positive definite. In particular, all the diagonal entries are positive

Conditioning of a problem :-

Conditioning pertains to the perturbation behaviour of a mathematical problem. Stability pertains to the perturbation behaviour of an algorithm used to solve that problem on a computer.

We can view a problem as a function $f: X \longrightarrow Y$ from a normed vector space X of data to normed vector space Y of solutions. This function f is usually non linear, but most of the time it is at least continuous.

A well conditioned problem is one with the property that all small purturbations of x lead to only small changes in $f(x)$.

An ill conditioned problem is the one with the property that some small purterbations of x leads to a large change in $f(x)$.

Absolute Condition Number :-

Let δx denotes a small perturbation of x and write $\delta f = f(x+\delta x) - f(x)$. The absolute condition number $\hat{k} = \hat{k}(x)$ of the problem f at x is defined as.

$$\hat{k} = \lim_{\delta \to 0} \sup_{\|\delta x\| \le \delta} \frac{\|\delta f\|}{\|\delta x\|}$$

the lim of sup in this formula can be interpreted as a sup over all infinitesimal perturbations δx and in interest of readibility,

$$\hat{k} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|}$$ δx and δf are infinitisimal.

The definition of the derivative gives us to the first order,

$\delta f \approx J(x)\,\delta x. \qquad \|\delta x\| \to 0.$

$$\hat{k} = \|J(x)\|$$

$J(x)$ be the matrix whose i, j entry is the partial derivative $\frac{\partial f_i}{\partial x_j}$ evaluated at x.

$\|J(x)\|$ represents the norm $J(x)$ induced by the norms on X and Y.

Relative condition Number :-

$$k = k(x)$$

$$k = \lim_{\delta \to 0} \sup_{\|\delta x\| < \delta} \left(\frac{\|\delta f\|}{\|f(x)\|} \Big/ \frac{\|\delta x\|}{\|x\|} \right)$$

If f is differentiable,

$$k = \frac{\|J(x)\|}{\|f(x)\| / \|x\|}$$

→ more important in numerical analysis

Examples :- T.B Book Page :- 91

$f(x) = x_1 - x_2$ is ill-conditioned.

Polynomial root finding is typically ill-conditioned even in cases that do not involve multiple roots.

Condition of Matrix-vector Multiplication :-

fix $A \in \mathbb{C}^{m\times n}$ and consider the problem of computing Ax from i/p x. That is we are going to determine a condition number corresponding to purterbations of x but not A.

$$K = \sup_{\delta x} \frac{\|A\delta x\|}{\|\delta x\|} \Big/ \frac{\|Ax\|}{\|x\|}$$

$$K = \|A\| \cdot \frac{\|x\|}{\|Ax\|}$$

A → square and non singular.

$$\frac{\|x\|}{\|Ax\|} \leq \|A^{-1}\| .$$ use use this to loosen a to a bound

independent of x:

$$K \leq \|A\|\|A^{-1}\|$$

$$K = \alpha \|A\| \|A^{-1}\|$$

- Stability of common operations:

• Addition, subtraction, multiplication, division.	Backward stable
• Inner product $x^* y$	Backward stable
• Outer product xy^*	Forward stable, not backward stable.
• Matrix-matrix multiplication	(Think about it / experiment)
• GE without pivoting	Unstable
• GE with PP	In next review.
• GE with CP	Backward stable
• Back substitution, forward substitution	Backward stable.
• E-value computation as roots of char. poly.	Unstable.

Special Linear systems.

diagonally dominant :- $A = LU$.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & d-\left(\frac{c}{a}\right)b \end{bmatrix}$$

If a & d ~~dom~~ 'dominate' b & c in magnitude then elements of L & U will be bounded. This will make LU factorization stable.

① **Row diagonally dominant :-** $A \in \mathbb{R}^{n \times n}$

$$\text{if } |a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \quad \forall i \in \{1, 2, \dots n\}$$

② **column diagonally dominant :-**

$$\text{if } |a_{ii}| \geq \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \quad \forall j \in \{1, 2, \dots, n\}$$

Theorem :-
If A is non singular & CDD, then it has a LU factorization & the entries of $L = (l_{ij})$ satisfy $|l_{ij}| \leq 1$

Theorem :- If $A \in \mathbb{R}^{n \times n}$ is symmetric & the principal submatrix $A(1:k, 1:k)$ is nonsingular for $k = 1, \dots, n$, then $\exists$ a lower Δ^r matrix L & $D = \text{diag}(d_1, \dots, d_n)$ such that $A = LDL^T$; the factorization is unique.

Theorem :-

If $A \in \mathbb{R}^{n\times n}$ is positive definite & if $X \in \mathbb{R}^{n\times m}$ is full rank, then $X^T A X \in \mathbb{R}^{m\times m}$ is also +ve definite

Cor :- If A is p.d then all its submatrices are p.d

Theorem :- A is pd $\Leftrightarrow$ $T = \frac{A+A^T}{2}$ is pd.

Cor :- If A is pd, then A has LU factorization & the diagonal entries of U are positive

Q When it is safe to avoid pivoting in positive definite system?

\# $A = T + S$ A is positive definite,
if $\Omega = \frac{\|S T^{-1} S\|_2}{\|A\|_2}$ is not too large, then it is ~~safte~~ safe to avoid pivoting

\# When A is symmetric then $\Omega = 0$, so, in this case LU factor of A exists & is stable to compute.

Cholesky factorization :-

If $A \in \mathbb{R}^{n\times n}$ is symmetric p.d. then $\exists$ a unique lower triangular $G \in \mathbb{R}^{n\times n}$ with positive diagonal entries such that $A = GG^T$.

G is called cholesky factor of A

For symmetric positive definite matrices the eigen values & singular values are the same.

To avoid issues arising from small e-values ie ill-conditionedness of A in this case, a factorization procedure with symmetric pivoting is preferred.

Overview of projection matrixe -

A projection matrix is a matrix P if $P^2 = P$.

Properties :-

① If $v \in \text{range}(P)$ $\quad Pv = v$.

② If $v \notin \text{range}(P)$ then $P(Pv - v) = 0$

$\Rightarrow Pv - v \in \text{null}(P)$

③ If P is a projection then $I-P$ is also projection.

④ $\text{range}(I-P) = \text{null}(P)$

⑤ $\text{range}(P) = \text{null}(I-P)$

⑥ $\text{range}(P) \cap \text{null}(P) = \{0\}$

An orthogonal projection P is one for which

$\text{range}(P) \perp \text{null}(P)$.

orthogonal projections are not orthogonal matrices

Iterative method :-

$Ax = b$.

$A = S - T$

S is non singular.

$Ax = (S-T)x = b$.

or, $x = S^{-1}Tx + S^{-1}b$

$$\boxed{R = S^{-1}T \quad , c = S^{-1}b.}$$

$x = Rx + c$

$x_{k+1} = Rx_k + c$

If $\|R\| < 1$, then $x_{k+1} = Rx_k + c$ converges for any x_0.

Spectral radius :- $\rho(R)$ = largest eigen value of R in abs. value.

① $\rho(R) \leq \|R\|$

② for every $\epsilon > 0$ $\exists$ $\|\cdot\|_\infty$ such that $\|R\|_\infty \leq \rho(R) + \epsilon$.

Theorem :- The iterative method $x_{k+1} = Rx_k + c$ converges to the solution of $Ax = b$, $\forall x_0$ & $\forall b$. $\Leftrightarrow$ $\boxed{\rho(R) < 1}$

① Jacobi

② Gauss Seidel

③ SOR

Jacobi :- $A = \cancel{D-E-F}$ $D-E-F$

$D =$ diag

$-E =$ lower triangular of A

$-F =$ upper triangular of A

$x_{k+1} = Jx_k + c \qquad c = D^{-1}b.$

$\downarrow$

Jacobi matrix

$J = I - D^{-1}A$

Gauss seidel :- The number of allocations required in Jacobi method can be reduced using the following trick :- use x_1^{k+1} to calculate x_2^{k+1} and so on.

Successive over-relaxation :- ω - relaxation parameter.

idea :- to improve on G-S by taking appropriate weighted average of x_j^{k+1} & x_i^k s.

$$A = \left(\frac{D}{\omega} - E\right) - \left(\frac{1-\omega}{\omega} D + F\right)$$

$$x_{k+1} = \left(\frac{D}{\omega} - E\right)^{-1} \left(\frac{1-\omega}{\omega} D + F\right) x_k + c$$

If $\omega > 1$, the method is called 'over relaxation'

$\omega < 1$ 'under relaxation'.

Key facts about Jacobi's method -

1. requires non-zero diagonal entries (can be accomplished by permuting rows/columns if not already true)
2. requires $2n$ memory allocations ($\dim A = n \times n$).
3. components do not depend on one another, so they can be computed simultaneously.
4. does not always converge; converges for sure when A is SDD. $\left[\text{i.e. } |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right]$

Key facts about G-S method.

1. requires non-zero diagonal entries.
2. requires n memory allocations at each step.
3. each component depends on previous ones, so must be computed successively.
4. converges if A is HPD (weaker than SDD)
5. When they converge together, G-S cges twice as fast as Jacobi.

<u>Key facts/questions about SOR</u>:

1. requires non-zero diagonal entries.
2. 2 questions: (i) for what ω is $\rho(L_\omega) < 1$?
   Is there an interval $I \subset \mathbb{R}$ such that $\rho(L_\omega) < 1 \;\; \forall \omega \in I$.

   (ii) Is there an optimal ω_0 such that

$$\rho(L_{\omega_0}) = \inf_{\omega \in I} \rho(L_\omega).$$

$x^{k+1} = L_\omega x^k + c$

By choosing appropriate ω it is possible to attain a higher rate of convergence than G-S.

rate of convergence

$$r(R) = -\log_{10} \rho(R)$$

Least Square Problems :-

If A has full rank, then the solution x to the least square problem is unique and is given by

$Ax = Pb$ i.e $x = \underbrace{(A^* A)^{-1} A^*}_{\text{Pseudo inverse of A denoted by } A^+} b.$

The LSP for a full rank matrix A reduces to computing $x = A^+ b$.

methods :-

① Solve the normal Equations i.e. $A^* A x = A^* b$.

Solved by cholesky operation count $\underbrace{mn^2}_{A^*A} + \frac{n^3}{3}$

② Using QR :- householders

flop count $2mn^2 - \frac{2n^3}{3}$ flops.

③ using SVD

2 norm is good because

① $\phi(x) = \frac{1}{2}\|Ax - b\|_2^2$ is a differentiable function.

② 2 norm is invariant under unitary / orthogonal transformations.

$Ax = b.$

$Q^T A x = Q^T b.$

for some orthogonal matrix Q.

A is full rank, trouble can be expected if A is 'nearly rank deficient' i.e columns of A are nearly independent.

Rank deficient LSPs.

① Infinitely many solutions.

First numerical rank of A must be determined & then the solution can be identified

To solve it we use complete orthogonal factorizations.

The 'SVD' is a particularly 'revealing' complete orthogonal factorization.

Eigen value Problems :-

- In general algebric multiplicity $\geq$ Geometric multiplicity
- If $\exists \lambda$ for which alg. multiplicity $>$ Geo. multiplicity. λ is 'defective eigen-value'
- A matrix with one or more defective eigen values is called a 'defective matrix'.
- A is non defective $\Leftrightarrow$ A has an evalue decomposition $X \Lambda X^{-1}$.

$\uparrow$ e-value revealing factorization.

Eigen value revealing factorization :-

① unitary diagonalization :-

If the e-vectors of A form an orthogonal basis of the underlying vector space, then the unitary matrix Q, whose columns are these e-vectors of A diagonalizes A. ie.

$$A = Q \Lambda Q^{*}.$$

A is unitary diagonalizable

$\Downarrow$

It is $\boxed{\text{normal}}$

If A is unitary diagonalizable it is normal.

② Schur factorization :-

$A = QTQ^{*}$ $\quad$ Q is unitary
$T =$ upper triangular

e-values of A appear on the diagonal of T since A and T are similar.

Theorem :-

Every square matrix $A \in \mathbb{C}^{m\times m}$ has a schur factorization.

every square matrix over $\mathbb{C}$ is triangulable.

Q:- Why can we not have direct methods for e-value computations?

① Finding e-values of a $m\times m$ matrix.

② finding roots of a poly. of degree m.

Abel's theorem → there doesn't exist a formula for finding roots of an arbitrary polynomial of degree >5 .

direct & iterative :-

① direct method :-
- General use algo.
  - power iteration
  - Inverse iteration.
  - orthogonal Iteration
  - QR Iteration

• Two phase method & reduction to Hessenberg form.

• Algorithms for symmetric matrices.
  - Rayleigh quotient iteration
  - Jacobi method
  - Bisection method
  - Divide and conquer.

① Iterative :
① Arnoldi
② GMRES.
③ LANCZOS.

Power iteration:- This method finds the abs value of the largest e-value of A & its corresponding e-vector.

① x_0 is chosen at random it must be non zero.

② The rate of convergence is completely dependent on the gap between λ_1 & λ_2 (eigen gap)

If $\lambda_1 >> \lambda_2$ convergence is fast

$\frac{\lambda_2}{\lambda_1} \approx 1$ the convergence is very slow

③ If A is real & the largest e-value is complex then there are 2 complex-conjugate e-values with the same abs value; then this method doesn't work.

Inverse iteration:-

useful when we know for an e-value λ and we want to find the corr. evector.

- Apply power method to a shift $(A - \sigma I)^{-1}$
- This makes the power method converge to an e-value closest to σ.

Orthogonal Iteration :-

simultaneous.	apply the power iteration
block power.	to several vectors at once.
subspace	Instead of starting with one
	vector x_0, suppose we start with
	m L.I ~~ve~~ & orthogonal vectors.

$$|\lambda_1| > |\lambda_2| \cdots > |\lambda_m|$$

QR Iteration :-

Assumptions.

$$|\lambda_1| > \cdots > |\lambda_m|$$

Z_0 = initial matrix consisting of m orthogonal colums.