cheatsheet

l1 norm: vector -> $\sum_{i=1}^{n}|x_i|$, matrix -> $\max_j ||a_j||_1$ (max absolute column sum)
l2 norm: vector -> $||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$, matrix ->largest singular value of A
orthogonal matrixes preserve the ℓ_2 norm when multiplied
l∞ norm: vector -> $||x||_\infty = \max_{1\le i\le n}|x_i|$ matrix -> max absolute row sum
general norm definiton

- positive definiteness $||x|| \geq 0 \wedge ||x|| = 0 \iff x = 0$
- scalar multiplication $||\alpha x|| = |\alpha|\ |x||$
- triangle inequality $||x+y|| \leq ||x|| + ||y||$
- matrix: +sub-multiplicaitve
  metric space
- $d(x,y) \geq 0$, $d(x,y) = 0 \iff x = y$, $d(x,z) \leq d(x,y) + d(y,z)$,$d(x,y) = d(y,x)$
  triangular inequality: $\forall A, B \in \mathbb{R}^{m\times n}, ||A+B|| \leq ||A|| + ||B||$
  sub-multiplicative $\forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}, ||AB|| \leq ||A||\ ||B||$
  subordinate matrix norm for A: $||A|| = \sup\{||Ax|| : x \in \mathbb{R}^n, ||x|| = 1\}$ a subordinate implies $||M\vec{v}|| \leq ||M||||\vec{v}||$
  inner product: $< v, w >:= \sum_{i=1}^{n} v_i^* w_i$
  complex conjugate: $c = c_1 + ic_2, c^* = c_1 - ic_2$
  norm of $v \in \mathbb{C}^n$ $||v|| := \sqrt{< v, v >}$
  Least square method: first test $Ax = b$ has no solutions, then solve for the normal equations $A^T Ax = A^T b$
  the proof is done by constructing $b_r + b_n = A$ and $b_r \in range(A)$ $b_n \in null(A)$, then solve for minimum(find a square and $b_r b_n = 0$)
  algebraic multiplicity: characteristic function $(\lambda + a)^n(\lambda + b)^m$, m and n are the multiplicity
  geometric multiplicity: find the span, the number of vectors in the span
  spectral decomposition: $Q^T AQ = S$ S is diag(eigenvalues), Q is orthogonal $Q^T = Q^{-1}$ and is the combination of eigenvectors A is symmetric
  Positive definite: $\text{symmetric} A \in \mathbb{R}^{n\times n}, \forall x \in \mathbb{R}^n \setminus \{0\}. x^T Ax > 0$ or all eigenvalues positive
  Positive semi-definite: $\text{symmetric} A \in \mathbb{R}^{n\times n}, \forall x \in \mathbb{R}^n. x^T Ax \geq 0$ or all eigenvalue non-negative
  Singular value decomposition: $A = U \sum V^T$
- 1. m > n, if $A \in \mathbb{R}^{m\times n}$, get eigenvalues σ_i^2/vectors of $A^T A$ else AA^T
- 2. combine the normalised to get V, V should be $n \times n$
- 3. $u_i = \frac{1}{\sigma_1} Av$, if $\sigma_i = 0$, or no more v available, take a vector that is orthogonal to all the rest(cross product)
- 4. $\sum = diag(\sigma)$
- $||A||_2 = \max_i \sigma_i$
  non-singular: columns are linearly independent, otherwise singular, singular means non-invertible
  orthogonal matrix: $A^{-1} = A^T$
  $\sum \text{eigenvalues} = trace(A) = \sum A_{ii}$
  Jordon Normal Form the diagonal is full of eigenvalues of A and the superdiagonal(the diagonal just above) is all 1
  generalised eigenvector: $(A - \lambda I)^k w = 0$, or $(A - \lambda I)^k w = \text{some eigenvector}$
  similar matrices: non singular A, A and SAS^{-1} have the same set of eigenvalues and if v is a eigenvector of A, then Sv is a eigenvector of SAS^{-1}
  Lower(Upper) triangular matrices: $A_{ij} = 0 \forall j > i (i > j)$
  Cholesky factorisation: $A = LL^T$ symmetric A, lower triangular L, then $Ax = b \Rightarrow LL^T x = b \Rightarrow y = L^T x$, iff A is positive definite
  projection: $proj_u(v) = \frac{u \bullet v}{u \bullet u}$ if $u \neq 0$
  Gram Schmidt: $u_1 := v_1, u_n = v_n - \sum_{i=1}^{n-1} proj_{u_i}(v_n) = v_n - \sum_{i=1}^{n-1}(e_1 \bullet v_2)e_1, e_n = u_n/||u_n||$**GS and QR**: let u_i be the columns of A, linearly independent, then $Q := [e_1, \ldots, e_n] \in \mathbb{R}^{m\times n}$ is orthogonal and $R_{ij} = e_i \bullet a_j$
  Householder map: reflection of $\vec{x}$ through $P = \{x \in \mathbb{R}^m : u \bullet x = 0\}$,transformation matrix $H_u = I - 2uu^T$
  Cauchy sequence in metric space: (S, d) is a metric space, then sequence (a_n) in S $\forall \epsilon > 0, \exists N \in \mathbb{N}$ s.t. $\forall n, m > N, d(a_n, a_m) < \epsilon$
  complete: metric space is complete if every Cauchy sequence in the space is convergent, $(\mathbb{R}^n, || \bullet ||_p)$ is complete for $0 < p \leq \infty$,and also $(C[a,b], || \bullet ||_\infty)$
  contracting map: $f : X_1 \to X_2$ of metric spaces (X_1, d_1) and (X_2, d_2) is contracting if for some $\alpha \in [0,1)$ $d_2(f(x), f(y)) \leq \alpha d_1(x,y), \forall x, y \in X_1$
  Banach fixed point theorem: A contracting map $f : X \to X$
  on a complete metric space X has a unique fixed point given by $\lim_{k\to\infty} f^k(x_0)$ for any point $x_0 \in X$
  hermitian conjugate / adjoint A^*, transpose, and make every element the complex conjugate
  hermitian: $A = A^*$, symmetric in real
  theorem: P is an orthogonal projector if and only $P = P^*$
  classical GS $q_j = \frac{u_j}{||u_j||}, u_j = a_j - \sum_{i=1}^{j}(u_j^* a_j)u_i$

Algorithm CGS
for j = 1 **to** n **do**
  $u_j = a_j$
  for i = 1 **to** j - 1 **do**
    $r_{ij} = q_i^* a_j$
    $u_j = u_j - r_{ij} q_i$
  end for
  $r_{jj} = ||u_j||_2$
  $q_j = u_j / r_{jj}$
end for

modified GS $q_j = \frac{P_j a_j}{||P_j a_j||}, P_j = I - \hat{Q}_{j-1}\hat{Q}_{j-1}^*, \hat{Q}_{j-1} = [q_1| \ldots |q_{j-1}]$

Algorithm original MGS
for j = 1 **to** n **do**
  $u_j = a_j$
end for
for j = 1 **to** n **do**
  $r_j j = ||u_j||_2$
  $q_j = u_j / r_{jj}$
  for k = j+1 **to** n **do**
    $r_{jk} = q_j^* u_k$
    $u_k = u_k - r_{jk} q_j$
  end for
end for

problem: $f : \mathcal{X} \to \mathcal{Y}$
well/ill-conditioned: small perturbations in x produce small/large changes in $f(x)$
conditional number: $cond(x) = \max_\delta \frac{||f(x) - f(x+\delta)||}{||\delta||}$ or $cond(x) = \lim_{\delta\to 0} \max_{||\delta x|| \leq \delta} \frac{||f(x) - f(x+\delta x)||}{||\delta x||}$
for matrix:

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

relative condition number: $\kappa(x) = \max_{\delta x}(\frac{\frac{||\delta f||}{||f(x)||}}{\frac{||\delta x||}{||x||}})$
if f differentiable: $\kappa(x) = \frac{||J(x)||}{||f(x)||/||x||}$
algorithm: $\tilde{f} : \mathcal{X} \to \mathcal{Y}$
stable: $\frac{\tilde{f}(x) - f(\tilde{x})}{||f(\tilde{x})||} = \epsilon_{\text{machine}}; \frac{||\tilde{x} - x||}{||x||} = \mathcal{O}(\epsilon_{\text{machine}})$
backwards stable $\tilde{f}(x) = f(\tilde{x}); \frac{||\tilde{x} - x||}{||x||} = \mathcal{O}(\epsilon_{\text{machine}})$
Gaussian elimination $Ax = b \Rightarrow A = LU$
$L = L_1^{-1} L_2^{-1} \ldots L_{m-1}^{-1}$ where L_n is the matrix from row operations, if the pivot is in column n and $Row\ m := Row\ m - a \bullet Row\ 1$, then $L_{nm} = a$

Algorithm GE
$U = A, L = I$
for k = 1 **to** m-1 **do**
  for j = k+1 **to** m **do**
    $\ell_{j,k} = u_{j,k} / u_{k,k}$
    $u_{j,k:m} = u_{j,k:m} - \ell_{j,k} u_{k,k:m}$
  end for
end for

partial pivoting $L_{m-1}P_{m-1} \ldots L_2 P_2 L_1 P_1 A = U$
before each step, we find the none final-state row with the largest pivot, and swap it to the top

Algorithm GE
$U = A, L = I, PI$
for k = 1 **to** m-1 **do**
  $\arg\max_{i\geq k} |u_{i,k}|$
  swap $u_{k,k:m}, u_{i,k:m}$
  swap $\ell_{k,1:k-1}, \ell_{i,1:k-1}$
  $p_{k,:}, p_{i,:}$
  for j = k+1 **to** m **do**
    $\ell_{j,k} = u_{j,k} / u_{k,k}$
    $u_{j,k:m} = u_{j,k:m} - \ell_{j,k} u_{k,k:m}$
  end for
end for

stability analysis, growth factor $\rho = \frac{\max_{i,j}|u_j|}{\max_{i,j}|a_{i,j}|}$
it is backward stable if $\rho = \mathcal{O}(1)$
GE with pivoting has $\rho \leq 2^{m-1}$
iterative methods for linear systems: $Ax = b \Rightarrow x^{(k+1)} = Bx^{(k)} + d, k = 0, 1, 2, \ldots$
stop criterion $\frac{||b - Ax^{(k)}||}{||b||} \leq \epsilon$
iterative methods:

- Jacobi iteration $x_i^{(k+1)} = \frac{1}{a_{i,i}}\left(b_i - \sum_{j=1, i\neq j}^{n} a_{i,j} x_j^{(k)}\right)$
- Gaussian-Seidel iteration $x_i^{(k+1)} = \frac{1}{a_{i,i}}\left(b_i - \sum_{j=1}^{i-1} a_{i,j} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j} x_j^{(k)}\right)$
- Successive over relaxation (SOR) iteration:

- $x_i^{(k+1)} = \frac{\omega}{a_{i,i}}\left(b_i - \sum_{j=1}^{i-1} a_{i,j}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j}x_j^{(k)}\right)$
- **Convergence of** $x^{(k+1)} = Bx^{(k)} + d$: if $||B|| < 1$, then the sequence converges for any starting point $x^{(0)}$

  Convergence if A is strictly row-diagonally dominant ($|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|$), then Gauss-Seidel and Jacobi converge

  Convergence if A is symmetric positive definite, then Gauss-Seidel and SOR converge ($\omega \in (0,2)$)

  partial derivatives: $f_{x_i}(\vec{x}) = f_{x_i} = \frac{\partial}{\partial x_i} f(\vec{x}) = \frac{\partial f}{\partial x_i} = \lim_{\delta\to 0} \frac{f(\vec{x}+\delta e_i)-f(\vec{x})}{\delta}$

  Clairaut's theorem: suppose f is defined over $\mathcal{D}$ and that $f_{x_i x_j}$ and $f_{x_j x_i}$ are both continuous on $\mathcal{D}$. Then for $\vec{x} \in \mathcal{D}$ $f_{x_i x_j}(\vec{x}) = f_{x_j x_i}(\vec{x})$

  ∇: $D_u f(\vec{x}) = \nabla f(\vec{x}) \bullet u = |\nabla f(\vec{x})||u| \cos\theta$, $\nabla^2 f = H$(hessian)

  local minimum: $\nabla f(\vec{x}) = 0$, $\nabla^2 f(\vec{x})$ is positive definite

  gradient descent: $Ax = b \iff \min_x f(x) = \frac{1}{2}x^t Ax - x^T b$, $\nabla f(x) = Ax - b, \nabla^2 f(x) = A$

  conjugate gradient

$$k = 0 : p^{(0)} = -\nabla f(x^{(0)}) = b - Ax^{(0)} = r^{(0)}$$
$$k \geq 1 : p^{(k)} = r^{(k)} - \sum_{i<k} \frac{{p^{(i)}}^T A r^{(k)}}{{p^{(i)}}^T A p^{(i)}} p^{(i)}$$
$$\alpha^{(k)} = \arg\min_{\alpha} f(x^{(k)} + \alpha^{(k)} p^{(k)}) = \frac{{p^{(i)}}^T A r^{(k)}}{{p^{(i)}}^T A p^{(i)}}$$

  without rounding errors, CG converges in $\leq m$ iteration, residual vectors are orthogonal

  power iteration

$$\begin{aligned} x^{(0)} &= a_1 q_1 + a_2 q_2 + \cdots + a_m q_m \\ x^{(k)} &= x_k A^k x^{(0)} = c_k(a_1\lambda_1^k q_1 + a_2\lambda_2^k q_2 + \cdots + a_m \lambda^m q_m) \\ &= c_k \lambda_1^k (a_1 q_1 + a_2(\lambda_1/\lambda_2)^k q_2 + \cdots + a_m(\lambda_m/\lambda_1)^k q_m) \end{aligned}$$

Algorithm Power iteration

```
for k = 1, 2, 3, ... do
    x̂^(k) = A x^(k-1)
    x^(k) = x̂^(k) / max(x̂^(k))
    λ^(k) = (x^(k))^T A x^(k)
end for
```

Algorithm Inverse iteration

```
for k = 1, 2, 3, ... do
    x̂^(k) = (A - σI)^(-1) x^(k-1)
    x^(k) = x̂^(k) / max(x̂^(k))
    λ^(k) = (x^(k))^T A x^(k)
end for
```

Algorithm Rayleigh quotient iteration

```
for k = 1, 2, 3, ... do
    x̂^(k) = (A - λ^(k-1) I)^(-1) x^(k-1)
    x^(k) = x̂^(k) / max(x̂^(k))
    λ^(k) = (x^(k))^T A x^(k)
end for
```

Rayleigh quotient: symmetric square matrix A, and vector x$r(x) = \frac{x^T Ax}{x^T x}$

Algorithm Basic QR iteration

```
for k = 1,2,3 do do
    A^(k-1) = Q^(k-1) R^(k-1)
    A^(k) = R^(k-1) Q^(k-1)
end for
```

$$\begin{aligned} A^{(1)} = R^{(0)}Q^{(0)} &= (Q^{(0)})^T A^{(0)} Q^{(0)} \\ A^{(2)} = R^{(1)}Q^{(1)} &= (Q^{(1)})^T A^{(1)} Q^{(1)} \end{aligned}$$