Orthogonality

- <u>inner product</u> (dot product): the inner product of two vectors x and y in $\mathbb{R}^n$ is

$$< x,y > := x \cdot y = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

1. we can express it in matrix notation

$$\langle x,y \rangle = x^T y = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

2. $\langle x,y \rangle = \langle y,x \rangle$ (only true for real numbers)
3. $\langle x, cy + dz \rangle = c\langle x,y \rangle + d\langle x,z \rangle$ where $c,d \in \mathbb{R}$ and $x,y,z \in \mathbb{R}^n$
4. $\langle \mathbf{x}, \mathbf{Ay} \rangle = \langle \mathbf{A^T x}, \mathbf{y} \rangle$
5. can write the 2-norm as inner product

$$\langle x,x \rangle = \sum_{i=1}^{n} x_i^2 = \|x\|_2^2$$

6. $|\langle x,y \rangle| \leq \|x\|_2 \cdot \|y\|_2$ (Cauchy–Schwarz inequality)
7. $\langle x,y \rangle = \|x\|_2 \cdot \|y\|_2 \cdot \cos\theta$ where θ is angle b/t x & y

- <u>orthogonal **vectors**</u>: two vectors are orthogonal if $\langle x,y \rangle = 0$ $(x \perp y)$
- <u>orthogonal **sets**</u>: vector set $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ are said to be orthogonal if $< x_i, x_j >= 0 \ \forall i \neq j$
- <u>orthonormal **sets**</u>: they are orthogonal & the vectors are unit vectors

$$< x_i, x_j > = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} =: \delta_{ij}$$

- <u>orthogonal **subspace**</u>: 2 subspaces S_1 and S_2 are orthogonal iff

$$\forall u \in S_1, \forall v \in S_2, < u,v >= 0$$

  ○ theorem: $S_1 \perp S_2$ iff there exists a basis $B = \{b_1, \ldots, b_k\}$ for S_1 and $C = \{c_1, \ldots, c_l\}$ for S_2 that's mutually orthogonal; i.e

$$< b_i, c_j >= 0 \qquad \forall i = 1, \ldots, k, \quad \forall j = 1, \ldots, l$$

  ○ note: if one such pair of basis exist, then any basis for each S_i also satisfy this property (so enough to find a basis pair and check)
  ○ note: the property above is equivalent to

$$B^T C = \vec{0} \qquad \text{where } B = [b_1 | \ldots | b_k] \ \& \ C = [c_1 | \ldots | c_l]$$

- <u>orthogonal complement</u>: let U be a subspace of W, we define the orthogonal complement of U as

$$U^\perp = \{x \in W : x \perp U\}$$

  ○ (all vectors that are orthogonal to every vector in U)
  ○ note: $U^\perp$ is the largest subspace that is orthogonal to U
  ○ $\dim(U) + \dim(U^\perp) = \dim(W)$ (they make up the entire space)
  ○ $(U^\perp)^\perp = U$

  ○ $\text{basis}(U) \cup \text{basis}(U^\perp) = \text{basis}(W)$
  ○ <u>orthogonal decomposition</u>: we can express any vector $x \in W$ as

$$x = x_u + x_{u^\perp} \qquad\qquad x_u \in U, \ x_{u^\perp} \in U^\perp$$

- important fact: let A be a $m \times n$ matrix, then
  1. $N(A) = [R(A^T)]^\perp$
  2. $N(A^T) = [R(A)]^\perp$

Orthogonal Projection

- <u>projection onto **vectors**</u>: projection of vector x onto vector u is

$$\text{proj}_u(x) = \frac{\langle x,u \rangle}{\langle u,u \rangle} u$$

$$\text{proj}_u(x) = \langle \hat{u}, x \rangle \cdot \hat{u} \qquad \hat{u} = \frac{u}{\|u\|} = \text{unit vector of } u$$

$$\text{proj}_u(x) = \frac{uu^T}{\|u\|^2} x = P_u x \qquad \therefore P_u = \frac{uu^T}{\|u\|^2}$$

  ○ call P_u the orthogonal projection matrix onto span u

- properties of P_u: let $P_u = P$ for notation purposes
  1. $P_u(P_u x) = P_u(x) \longrightarrow (P_u)^k = P_u$
     ○ can't project something out of a span once it's already in it
  2. $(P_u)^T = P_u$
  3. $R(P) = \text{span}\{u\}$ and $N(P) = \text{span}\{u\}^\perp$

- <u>orthonormal basis</u>: we say $\{w_1, w_2, \ldots, w_m\}$ is an orthonormal basis (ONB) for a subspace U if $\{w_1, \ldots, w_m\}$ is a basis for U and

$$\langle w_i, w_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad (\{w_1, \ldots, w_m\} \text{ is orthonormal})$$

- properties of ONB: let $\{w_1, w_2, \ldots, w_m\}$ be ONB for U, then for any $x \in U$, there exist a unique set of scalars $\{c_1, c_2, \ldots c_m\}$ s.t
  1. $x = \sum_{i=1}^m c_j w_j$ \qquad (not special, just definition of basis)
  2. $c_j = \langle w_j, x \rangle$ \qquad (special to ONB)
  3. $\|x\|^2 = \sum |c_j|^2$ \qquad (Parseval's Equality – holds for all basis)

- **Gram-Schmidt Orthogonalization Algorithm**: let $\{v_1, \ldots, v_n\}$ be basis of subspace U, we want to find the ONB

$$u_1 = v_1 \qquad\qquad u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1$$

$$u_3 = v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2$$

$$u_n = v_n - \sum_{j=1}^{n-1} \frac{\langle v_n, u_j \rangle}{\langle u_j, u_j \rangle} u_j = v_n - \sum_{j=1}^{n-1} P_{v_j}(u_n)$$

$$e_i = \frac{u_i}{\|u_i\|}$$

  ○ $\{u_1, u_2, \ldots, u_n\}$ is an <u>orthogonal</u> basis of U
  ○ $\{e_1, e_2, \ldots, e_n\}$ is an <u>orthonormal</u> basis of U (it's been normalized)

- <u>projection onto **subspaces**</u>: let $U \subseteq R^n$ be a subspace with ONB $\{w_1, \ldots, w_m\}$, then

$$\text{proj}_U(x) := \text{proj}_{w_1}(x) + \text{proj}_{w_2}(x) + \ldots \text{proj}_{w_m}(x)$$
$$= \left(w_1 w_1^T + w_2 w_2^T + \ldots w_m w_m^T \right) x$$
$$= P_U \times x$$

  ○ P is the ortho projector onto U (it is a matrix)
  ○ (second line works because w_j is a unit vector)
  ○ **important**: alternate way to find P_U

$$B = \begin{bmatrix} w_1 & | & w_2 & | & \ldots & | & w_m \end{bmatrix} \quad \text{(ONB as columns)}$$
$$P_U = BB^T$$

- <u>ortho-projector matrix</u>: a matrix P is an ortho projection matrix iff $P^2 = P$ and $P^T = P$
  ○ **fact**: if P is an ortho projector onto U, then $Q = I - P$ is the ortho projector onto $U^\perp$ (projects any vector onto $U^\perp$)
  ○ $x - P_u(x) \in U^\perp$ for any vector x
  ○ $\|x - P_u(x)\| \leq \|x - y\| \quad \forall \ y \in U$
    $\rightarrow$ basically saying the orthogonal projection of x onto U ($P_U(x)$) is the closest point in U to x
    $\rightarrow$ thus $\|x - P_U(x)\|$ is distance from x to the closest point in U

QR Decomposition

- <u>orthogonal matrix</u>: a matrix A is "orthogonal" if $A^T A = AA^T = I$
  1. A is square and invertible ($A^{-1} = A^T$)
  2. $\|Ax\| = \|x\|$ (norm preserving or has norm of 1)
  3. columns of A are orthonormal
  4. rows of A are orthonormal

  ○ ex. identity, rotation and reflection matrices are all orthogonal

- <u>reflection matrix</u>: reflection of vector x across subspace U is

$$\text{ref}_U(x) = (I - 2P_{U^\perp})x$$

  ○ for any ortho projector P, the reflection matrix is $I - 2P$ and it's also orthogonal

- **QR Decomposition**:
  1. Write $A = \begin{bmatrix} a_1 & | & \ldots & | & a_n \end{bmatrix}$
  2. Apply Gram-Schmidt to $\{a_1, \ldots, a_n\}$ and construct $\{w_1, \ldots, w_n\}$
  3. Rewrite each column a_j of A as linear combination of the ONB

$$A = Q_1 R_1 \qquad Q_1 = \underbrace{\begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}}_{m \times n}$$

$$R_1 = \underbrace{\begin{bmatrix} \langle w_1, a_1 \rangle & \langle w_1, a_2 \rangle & \ldots & \langle w_1, a_n \rangle \\ & \langle w_2, a_2 \rangle & \ldots & \langle w_2, a_n \rangle \\ & & \ddots & \\ & & & \langle w_n, a_n \rangle \end{bmatrix}}_{n \times n}$$

(this is called the **thin QR decomposition** of A)

4. Obtain the full QR decomposition of A by writing

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

where Q is an $(m \times m)$ orthogonal matrix and R is a $(m \times n)$ upper triangular matrix
- let $Q_2 = \begin{bmatrix} w_{n+1} & w_{n+2} & \dots & w_m \end{bmatrix}$
- since $R(A)^\perp = N(A^T)$, **we just solve $A^T w = 0$ for Q_2**
- **note**: can also do $Q_2 = N(Q_1^T)$, remember to normalize

- **theorem**: let $A = QR$ be the full QR decomposition of the matrix A and let $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$

1. the columns of Q_1 form ONB for $R(A)$
2. the columns of Q_2 form ONB for $R(A)^\perp$

$$\text{proj}_{R(A)}(x) = Q_1 Q_1^T x$$
$$Q_1 Q_1^T = \text{ortho projector onto } R(A)$$
$$\text{proj}_{R(A)^\perp} = Q_2 Q_2^T x$$
$$Q_2 Q_2^T = \text{ortho projector onto } R(A)^\perp$$

Least Squares Approximation

- the normal equation: let A be an $m \times n$ matrix with and $m > n$ and $\text{rank}(A) \geq n$ – the least squares approximation of the system $Ax \approx b$ is the solution of the system

$$A^T A x^\star = A^T b$$
$$x^\star = (A^T A)^{-1} A^T b$$

- the LSE (in our current set-up) always has a solution
- any solution u of LSE minimze $\|Au - b\|_2$
- if $A^T A$ is invertible then LSE has a unique sol (iff $\text{rank}(A) = n$)

- solving LSE using QR decomposition: assume same set-up and let $A = Q_1 R_1$ be the thin QR decomposition

$$R_1 x = Q_1^T y \qquad \text{(usually easier to solve)}$$
$$x_{LS} = R^{-1} Q_1^T y$$

further, the residual is given by

$$\|Ax - b\| = \|Q_2^T b\|$$

- fitting models to data: we have m points $\{(t_1, y_1), \dots, (t_m, y_m)\}$ and want to fit best-fit line (minimize SSE) of form $y = c_1 + c_2 t + c_3 t^2$

$$A = \begin{bmatrix} 1 & t_1 & (t_1)^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & (t_m)^2 \end{bmatrix} \qquad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

and we want to solve $Ac \approx y$ using the LSE
- we assume $m \geq n$ and the function $f_1, f_2, \dots, f_n$ are linearly independent (so $\text{rank}(A) = n$)
- our model won't fit the line perfectly and will be better conditioned (less-likely to overfit)

Eigenvalues

- eigenvalue/eigenvector pair: let A be an $n \times n$ matrix, a scalar $\lambda \in \mathbb{R}$ and a non-zero vector $v \in \mathbb{R}^n$ is called an eival/eigvec pair if $Av = \lambda v$
  - characteristic polynomial of A: $c_A(\lambda) = \det(A - \lambda I)$
- finding eigenvalues of A: it is the root of $c_A(\lambda)$ (solve $c_A(\lambda) = 0$)
- finding eigenvector given eigenvalue:

  solve for v: $\quad (A - \lambda_j I)v = 0$

  - any vector in basis of $N(A - \lambda_j I)$ is the corresp eigenvec to λ_j
  - define $E_{\lambda_j} := N(A - \lambda_j I)$ as the eigenspace of λ_j

- multiplicity of eigenvalue: say $c_A(\lambda) = (\lambda - \lambda_1)^2 (\lambda - \lambda_2)^3 (\lambda - \lambda_3)$ where $\lambda_1 \neq \lambda_2 \neq \lambda_3$
  - algebraic multiplicity of $\lambda_1, \lambda_2, \lambda_3$

  $$m_1 = 2 \qquad m_2 = 3 \qquad m_3 = 1$$

  - geometric multiplicity of $\lambda_1, \lambda_2, \lambda_3$

  $$d_j := \dim(E_{\lambda_j}) \qquad j = 1, 2, 3$$

  - when $d_j < m_j$, that's called a defective eigenvalue
  - **theorem**: there exists an eigenbasis (eigenvector span $\mathbb{R}^n$) corresponding to A if $d_j = m_j$ for each eigenvalue of A

Diagonalization

- setting for this section: assume A is $(n \times n)$ with
  - $\lambda_1, \lambda_2, \dots, \lambda_3$: eigenvalues of A
  - $\{v_1, v_2, \dots, v_n\}$: eigenbasis of A such that $Av_j = \lambda_j v_j$
- diagonalizability: matrix A is **diagonalizable** if there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$
  - **fact**: A is diagonalizable if it has n distinct eigenvec (eigenbasis)
  - if A is diagonalizable, we can make the eigenvectors columns of P and eigenvalues as diagonal entries of D

$$P = \begin{bmatrix} | & | & | & | \\ v_1 & v_2 & \dots & v_n \\ | & | & | & | \end{bmatrix} \qquad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- application of diagonalization: power of matrices – let $A = PDP^{-1}$

  $$A^k = PD^k P^{-1}$$

  - formula also hold for negative k if all eigenvalues are non-zero (i.e if A is diagonalizable and invertible)
  - note: $D^{-1} = 1/\lambda_j$ for all diagonal values

- symmetric matrix: a square matrix A is symmetric if $A^T = A$
  - **fact**: all eigenvalues of a real symmetric matrix A are real
  - **fact**: let A be a real symmetric matrix, and λ_1, λ_2 are distinct eigenvalues with respective eigenvectors $v_1, v_2 \to$ then $v_1 \perp v_2$

- spectral theorem: let A be a real symmetric matrix, then there exists an orthogonal matrix P and diagonal matrix D such that $A = PDP^T$
  - we say that A is orthogonally diagonalizable

- **important remark**: let A be any real $m \times n$ matrix, then we have

1. $A^T A$ and AA^T will both be real symmetric matrices
2. $A^T A$ and AA^T are orthgonally diagonalizable (Spectral theorem)

Singular Value Decomposition (SVD)

- singular value decomposition (SVD): let A be a $(m \times n)$ real matrix, then there exists an orthogonal matrix P, Q and a "diagonal" matrix Σ such that $A = P\Sigma Q^T$
  - Σ is a diagonal matrix with r non-zero diagonal entries
  - the values $\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0$ (they are ordered) are the non-zero singular values of A

- some propositions/observations

1. if λ is a non-zero eigenvalue of AA^T, λ is also the eigenvalue of $A^T A$ (other might have $\lambda = 0$ if dim are mismatched)
2. all eigenvalues of $A^T A$ and AA^T are non-negative
3. if λ is a non-zero eigenvalue of AA^T, then λ has the same level of repetition in $A^T A$ and AA^T

$$\dim(N(AA^T - \lambda I)) = \dim(N(A^T A - \lambda I))$$

- **SVD steps**: let A be $m \times n$ and real

1. Find singular value for Σ ($\mathbf{m \times n}$):
   (a) find eigenvalue of either $A^T A$ or AA^T, order them
   (b) set $\sigma_k = \sqrt{\lambda_k}$
2. Construct the matrix Q ($n \times n$)
   (a) find the corresponding eigenvectors of $A^T A$
      - if missing eigenvalues, assume remaining eigval are 0
   (b) set the normalized corresponding eigenvectors as columns

   $$Q = \begin{bmatrix} | & | & | & | \\ q_1 & q_2 & \dots & q_n \\ | & | & | & | \end{bmatrix}$$

3. Construct the matrix P ($m \times m$)
   (a) let p_k be the columns of P, then we can take

   $$p_k = \frac{1}{\sigma_k} A q_k$$

   this will give you the first r columns of P
   (b) for the remaining $m - r$ columns, complete $p_1, \dots p_m$ to an ONB (recall: thin QR to full QR, solve for $A^T w = 0$)

- application of SVD

1. $\|A\|_{op} = \sigma_1$ (the largest singular value of A)
2. $\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2)^{1/2}$
3. $\text{rank}(A) = r$ (number of non-zero singular value)
4. if A is $n \times n$ and invertible, then $A^{-1} = Q\Sigma^{-1}P^T$
   - Σ^{-1} is Σ with all diagonal entries flipped, $d_{ij} = 1/d_{ij}$
   - note: this is not a SVD of A^{-1} because columns of Σ is not ordered, if you reorder them (reverse the column order for all matrices), you'll get the SVD of A^{-1}
5. $\|A^{-1}\|_{op} = 1/\sigma_r$ (1 over smallest singular value)
6. $\text{cond}(A) = \|A\|_{op} \times \|A^{-1}\|_{op} = \dfrac{\sigma_1}{\sigma_r}$
7. assume $P = \begin{bmatrix} p_1 & \cdots & p_m \end{bmatrix}$ and $Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}$, then
   - $\{p_1, \dots, p_r\}$ is an orthonormal basis of $R(A)$
   - $\{p_{r+1}, \dots, p_m\}$ is an orthonormal basis of $N(A^T)$
   - $\{q_1, \dots, q_r\}$ is an orthonormal basis of $R(A^T)$
   - $\{q_{r+1}, \dots, q_n\}$ is an orthonormal basis of $N(A)$.

LU Decomposition

- it's a way to factorize A such that $A = LU$ where U is REF of A and L is the steps that got there
  - L is unit lower triangular matrix and U is upper triangular matrix
- type of question: given $A = LU$, solve $LUx = b$
  1. let $Ux = y$ and solve $Ly = b$ for y (normal augmented matrix)
  2. then solve $Ux = y$ for x
- if we have $A = LU$, we have some special facts
  - rank$(A) = $ rank(U); and if A is square, $\det(A) = \det(U)$
  - recall: if a matrix B is triangular, $\det(B)$ is the product of all its diagonal entries; also $\det(AB) = \det(A) \times \det(B)$

Norms and Condition Number

- norm assign magnitude (size) to vectors/matrices
- vector norm: a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a vector norm iff
  1. $\|x\| > 0, \ \forall \ x \in \mathbb{R}^n$
  2. $\|x\| = 0$ if and only if $x = \vec{0}$ (the zero-vector)
  3. $\|cx\| = |c| \cdot \|x\|, \ \forall \ c \in \mathbb{R}, \ \forall \ x \in \mathbb{R}^n$
  4. $\|x + y\| \leq \|x\| + \|y\|$ (known as the triangle inequality)
  (important for proof questions)
  - ex. $n = 1$: in $\mathbb{R}$ the absolute value of x, $|x|$ does the job
  - ex. $n = 2$: let $\vec{x} = <x_1, x_2>$, the typical norm is the Euclidean norm (aka 2-norm)

  $$\|x\| = \sqrt{|x_1^2| + |x_2^2|}$$

- matrix norm: a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a vector norm iff
  1. $\|A\| > 0$
  2. $\|A\| = 0$ iff A is a non-zero matrix
  3. $\|cA\| = |c|\|A\|$
  4. $\|A + B\| \leq \|A\| + \|B\|$
  5. $\|AB\| \leq \|A\| \times \|B\|$ (new)
  6. $\|Ax\|_2 \leq \|A\|_{op} \times \|x\|$

- operator norm: special kind of matrix norm - intuitively, it's calculating the maximum "stretching" capability of the matrix

  $$\|A\|_{op} = \max_{\|x\|=1} \|Ax\| = \text{max stretch of A}$$

  $$\|A^{-1}\|_{op} = \frac{1}{\min_{\|x\|=1} \|Ax\|_2} = \frac{1}{\text{min stretch of } A}$$

  - diagonal matrices: let D be a diagonal matrix, then the norm is the max magnitude of the diagonal entries

  $$\|D\| = \max\{|d_{jj}|\}$$

- condition number: helps tell us how "stable" a solution is (<u>lower is better</u>) $\to$ if A is invertible (nonsingular), then

  $$\text{cond}(A) = \|A\| \times \|A^{-1}\| = \frac{\text{max stretch of a unit vector}}{\text{min stretch of a unit vector}}$$

- relative error: given a vector b and a small change Δb, the relative error is defined as $\|\Delta b\|/\|b\|$
  - bound on relative error: if A is singular and we have $Ax = b$, say that some change Δb result in some change Δx, we have

  $$\frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(A)\frac{\|\Delta b\|}{\|b\|}$$

  - also we know that cond$(P) = 1$ for all permutation matrix P and so cond$(PA) = $ cond(A)

Interpolation

- given data $[(t_0, y_0), \ldots, (t_d, y_d)]$, an interpolating function (or interpolant) is a function $f(t)$ such that $f(t_k) = y_k$ for $k = 0, \ldots, d$
- polynomial interpolation: a polynomial of degree (at most) d is a function of the form

  $$p(t) = c_0 + c_1 t + \ldots c_d t^d, \quad c_i \in \mathbb{R}$$

  - note that there are $d + 1$ variables $\to$ want to solve for c_i
  - every data points give an equation - we want $p(t_j) = y_j$

  $$P(t_0) = c_0 + c_1 t_0 + \ldots + c_d (t_0)^d = y_0$$
  $$\ldots$$
  $$P(t_d) = c_0 + c_1 t_d + \ldots + c_d (t_d)^d = y_d$$

  put into matrix form and solve for vector $\vec{c} = [c_0, \ldots, c_d]^T$

  - note that with more data (large d) - the polynomial becomes very unstable (high condition number) and not useful

- cubic spline interpolation: consider $N + 1$ points $(t_0, y_0) \ldots, (t_N, y_N)$, a **cubic spline** is a function $p(t)$ defined piecewise (made up of many parts) by N cubic polynomials $p_1(t), \ldots, p_N(t)$ where

  $$p_k(t) = a_k(t - t_{k-1})^3 + b_k(t - t_{k-1})^2 + c_k(t - t_{k-1}) + d_k$$

  - need $4N$ equations - here first 2 give N each, next 2 give $N - 1$ each, last one give 2 equations

  $$p_k(t_{k-1}) = y_{k-1} \qquad p_k(t_k) = y_k$$
  $$p'_k(t_k) = p'_{k+1}(t_k) \qquad p''_k(t_k) = p''_{k+1}(t_k)$$
  $$p''_1(t_0) = p''_N(t_N) = 0$$

- given a coeff matrix C for a cubic spline, solve for certain missing cells

  $$C = \begin{bmatrix} 1 & -2 & 1 & a_4 & 1 & 1 \\ 0 & 3 & -3 & b_4 & -6 & -3 \\ 1 & 4 & 4 & c_4 & -5 & -14 \\ 1 & 3 & 8 & 10 & 9 & -1 \end{bmatrix}$$

  - <u>trick</u>: **iff** $t_k - t_{k-1} = 1$ for all points (i.e they're all 1 apart), then we can use these equations to find missing cells

  $$a_k + b_k + c_k + d_k = d_{k+1} \tag{1}$$
  $$3a_k + 2b_k + c_k = c_{k+1} \tag{2}$$
  $$6a_k + 2b_k = 2b_{k+1} \tag{3}$$

- another thing they ask is given the coefficient matrix C, find $p''(2.5)$
  - first need to find $p''_k(t) = 6a_k(t - t_{k-1}) + 2b_k$
  - for this you need to pick the right interval, $t = 2.5 \in [t_2, t_3]$ so you pick $p''_3(t)$ (always choose $k = $ right endpoint)

Subspaces

- subspace: a subset $S \subseteq \mathbb{R}^n$ is a **subspace** iff $\forall \ u, v \in S, \ \forall \ a \in \mathbb{R}$
  1. $u + v \in S$ (S is closed under addition)
  2. $a \times u \in S$ (S is closed under scalar multiplication)
- linear combination: the linear combination for a set of vectors $\{v_1, v_2, \ldots, v_k\} \in \mathbb{R}^n$ is the sum of it vectors with some scalar $c_j \in \mathbb{R}$ (c_j can be different for every j) i.e $v_{new} = \sum_{i=1}^k c_j v_j, \ c_j \in \mathbb{R}$
- span: the set of <u>all</u> linear combinations of $\{v_1, \ldots, v_k\}$ is its span

  $$\text{span}\{v_1, \ldots, v_k\} = \left\{\sum_{j=1}^k c_j v_j, \quad c_j \in \mathbb{R}\right\}$$

  (it is the set of all vectors that can be obtained by scaling and adding these vectors together)
  - for any $v_1, \ldots v_k \in \mathbb{R}^k$, span$\{v_1, \ldots, v_k\}$ is a subspace of $\mathbb{R}^k$
- linear dependence: $\{v_1, \ldots v_k\}$ is linearly dependent if at least 1 v_j can be expressed as linear combo of other vectors
  - if a set is not linearly dependent, it is <u>linearly independent</u>
  - <u>checking linear independence</u> of $\{v_1, \ldots, v_k\}$: put v_j as columns of a matrix, get to REF and check rank/pivots (vectors associated with pivot columns are linearly independent)
- basis vectors: a set $\{v_1, v_2, \ldots v_k\} \subseteq S$ is a basis of S if
  1. span$\{v_1, \ldots, v_k\} = S$ (the set of vector spans S)
  2. $\{v_1, \ldots, v_k\}$ is linearly independent
  - note: given a subspace, the choice of a basis is not unique
  - <u>dimension</u>: # of vector in any basis of subspace S (called dim(S))
  - <u>theorem</u>: in a k-dimension subspace S, any k linearly independent vectors $\{v_1, \ldots, v_k\} \in S$ form a basis for S
  - **important**: put spanning vectors into a matrix as columns, row reduce them, pivot columns tell you linearly independent vectors and those vectors is the basis of your span
- null space of A: it's defined as the solution set of $Ax = 0$

  $$N(A) = \{x \in \mathbb{R}^n : Ax = \vec{0}\} \subseteq \mathbb{R}^n$$

  - sample question: given A, find the basis of $N(A) \longrightarrow$ this means set up the augmented matrix and solve $Ax = 0$
- range of A: assume that A is $m \times n$

  $$R(A) := \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m = \text{span}\{a_1, \ldots, a_k\}$$
  $$= \text{set of all possible linear combinations of its columns}$$
  $$= \text{col}(A) = \text{"column space" of } A$$

  - sample question: given A, find the basis of $R(A)$
    $\to$ you again row reduce A and check for linear independence
    $\to$ the pivot columns of A indicates the linearly independent (original) columns of A, these make up the basis of $R(A)$

- important facts: let $A \in \mathbb{R}^{m \times n}$

  1. $\text{rank}(A) = $ # of pivots $\leq \min(m, n)$
  2. $\dim(R(A)) = \text{rank}(A)$ (b/c the # of lin indep col gives the basis)
  3. $\dim(N(A)) = $ # of free variable $= n - \text{rank}(A)$
  4. rank-nullity theorem: for any $m \times n$ matrix A

  $$\dim(R(A)) + \dim(N(A)) = n$$

- special case: if we have $A = LU$ and want to find $R(A)$ and $N(A)$

  - finding $R(A)$: let $A = LU$ and $\text{rank}(A) = r$, then the first r columns of L forms the basis for $R(A)$

  $$R(A) = \text{span}\{l_1, \ldots, l_r\}$$

  - finding $N(A)$: since L is invertible, null space of A is the same as the null space of U

  $$N(A) = N(LU) = N(U)$$

  so we just have to find $N(U)$ (means solve for $Ux = \vec{0}$)

    - $\rightarrow$ proposition: suppose B is invertible $m \times n$ and A is any $m \times n$ matrix, then $N(BA) = N(A)$ and $R(AB) = R(A)$
  - $R(A^T) = R(U^T) = $ the first r rows of U (not L)

Pseudoinverse

- **SVD expansion**: let A be a $m \times n$ matrix such that rank(A) = r and $A = P\Sigma Q^T$ is the SVD; then the SVD expansion of A is

  $$A = \sum_{k=1}^{r} \sigma_k p_k q_k^T$$

  where $p_1, \ldots, p_r$ are the first r columns of P (similar for q_i)

  - **fact**: the truncated SVD expansion of rank s is

  $$A_s = \sum_{k=1}^{s} \sigma_k p_k q_k^T$$

  - A_s is said to be a rank s approximation of A
  - A_s is the best rank s approximation wrt Frobenius norm

- **Principal Component Analysis** (PCA): given $x_1, x_2, \ldots x_n \in \mathbb{R}^P$, trying to find new set of variables (principal component) that capture the most significant variations in the data

  - (can assume data is centered (i.e $\sum x_k = 0$) but if not replace each points with $\tilde{x}_k = x_k - \bar{x}$)
  - we can form the data matrix X with x_i **as rows**
  - we want to find the unit vector w_1 that maximizes $\sum_{k=1}^{n} |\langle x_k, w_1 \rangle|$
  - (i.e 1st weight vector w_1 points in direction that captures most info (i.e max variance) of data, and 2nd weight vec $w_2 \perp w_1$
  - finding w_i: we can pick the weight vectors w_i as $w_i = q_i$ where q_i is the k-th column of Q in SVD decomposition of X

- **fact**: if A is $n \times n$ and invertible, then there's an $n \times n$ matrix such that $AA^{-1} = I$ and $A^{-1}A = I$ (i.e right inverse = left inverse)

- pseudoinverse: let A be an $m \times n$ matrix with SVD $A = P\Sigma Q^T$, we define the pseudoinverse $A^\dagger$

  $$A^\dagger = Q\Sigma^\dagger P^T$$

  - where $\Sigma^\dagger$ is Σ with $1/d_{ii}$ for non-zero diagonal entries

- **theorem**

  1. if A is invertible, $A^\dagger = A^{-1}$
  2. if A is $m \times n$, $m \leq n$ and $\text{rank}(A) = m$ then $AA^\dagger = I_m$ (R inverse)
  3. if A is $m \times n$, $n \leq m$ and $\text{rank}(A) = n$ then $A^\dagger A = I_n$ (L inverse)

- properties of pseudoinverse

  1. $AA^\dagger A = A$ and $A^\dagger AA^\dagger = A^\dagger$
  2. $AA^\dagger$ is the proj mat onto $R(A)$ and $A^\dagger A$ is proj mat onto $R(A^T)$
  3. let A be an $m \times n$ matrix with $\text{rank}(A) = n$ and let $b \in \mathbb{R}^m$, the LSE approximation of $Ax \approx b$ is given by

  $$x = A^\dagger b \qquad\qquad \therefore A^\dagger = \sum_{k=1}^{r} \frac{1}{\sigma_i} q_i p_i^T$$

  (can use this to solve LSE instead)

Complex Vectors

- trick to convert normal to polar : given $z = a + ib$, want $z = re^{i\theta}$
  - modulus $= r = |z| = \sqrt{a^2 + b^2}$ and $\theta = \cos^{-1}(a/r)$

- trick to convert polar to normal: given $z = re^{i\theta}$, want $z = a + ib$
  - $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ and multiply that by r

- complex vector space: a set of vectors of length n with complex entries
  - the conjugate of $v \in \mathbb{C}^n$ is the conjugate of each entry $\overline{v}_1, \ldots, \overline{v}_n$

- inner product: standard inner product of vectors $u, v \in \mathbb{C}^n$ is

  $$\langle u, v \rangle = u^T \overline{v} = u_1 \overline{v_1} + \ldots + u_n \overline{v_n}$$

  $$\langle cu, v \rangle = c\langle u, v \rangle \qquad \langle u, cv \rangle = \overline{c}\langle u, v \rangle \qquad \langle u, v \rangle = \overline{\langle v, u \rangle}$$

  - note: $\langle v, v \rangle \geq 0$ for all v and it's only 0 if $v = \vec{0}$

- conjugate transpose: the conjugate transpose of a complex matrices A is $A^\star = (\overline{A})^T$ (acts as a transpose but for complex matrices)
  - we can note that $\langle Au, v \rangle = \langle u, A^\star v \rangle$

- hermitian matrix: a complex matrix A is hermitian iff $A = A^\star$, they are **like symmetric matrices** and have similar properties

  1. $\langle Au, v \rangle = \langle u, Av \rangle$ for $u, v \in \mathbb{C}^n$
  2. A has real eigenvalues
  3. diagonal entries of A are real

- unitary matrix: a complex matrix A is unitary iff $A^{-1} = A^\star$ (check if $AA^\star = I$), they're like **orthogonal matrices** and have properties

  1. if A is real, then A is orthogonal
  2. $\langle Ax, Ay \rangle = \langle x, y \rangle$
  3. their columns and rows are orthonormal

- general spectral theorem: every hermitian matrix H is unitary diagonalizable with $H = UDU^\star$ (D is diagonal matrix with eigenvalues of H, U is matrix of eigenvectors of H, normalized to be orthonormal)

Fourier Basis

- note: moving forward we'll be using 0-indexing (like Python)

- roots of unity: an Nth root of unity is a complex number w s.t $w^N = 1$
  - there are N number of Nth root of unity (i.e $w^4 = 1$ has 4 solution)
  - let w_N be the (2nd) Nth roots of unity and z be set of all of them

  $$w_N = e^{2\pi i \cdot / N}$$
  $$z = \{(w_N)^k = e^{2\pi i \cdot k / N}, \quad k = 0, 1, \ldots, N - 1\}$$

  - properties
    1. $(w_N)^N = (w_N)^0 = 1$ (it repeats once you're past $N - 1$)
    2. $\overline{w}_N = (w_N)^{-1} = (w_N)^{N-1}$
       $\rightarrow$ means that $(1, N-1), (2, N-2)$, etc are conjugate pairs
    3. a full cycle of these roots sums up to 0

- Fourier basis: set of functions that span the space for period signals, any periodic signal can be rep as a combo of these basis functions
  - (similar to how any vector in R^n can be rep by standard basis)
  - let N be a positive integer and let $w_N = e^{2\pi i / N}$, the fourier basis of $\mathbb{C}^N$ is $\{f_0, \ldots, f_{N-1}\}$ where

  $$f_k = \begin{bmatrix} 1 & w_N{}^k & w_N{}^{2k} & \ldots & w_N{}^{(N-1)k} \end{bmatrix}^T$$

  (all the Nth root of unity raised to power of k)
  - properties
    1. Fourier basis is an **orthogonal basis** of $\mathbb{C}^N$
    2. $\|f_k\|_2 = \sqrt{N}$ (all of them have the same norm)
    3. $\overline{f_k} = f_{N-k}$ for $1 \leq k < N$ (or $\overline{f_k[i]} = f_k[N - i]$)
    4. if N is even, then $f_{N/2}$ is a real vector

Discrete Fourier Transform

- discrete Fourier transform: let $x \in \mathbb{C}^n$, then

  $$\text{DFT}(x) = F_N x = \begin{bmatrix} \overline{f_0}^T & \overline{f_1}^T & \ldots & \overline{f_{N-1}}^T \end{bmatrix}^T \cdot x$$
  $$= \begin{bmatrix} \langle x, f_0 \rangle & \langle x, f_1 \rangle & \ldots & \langle x, f_{N-1} \rangle \end{bmatrix}^T$$

  - note: the DFT is invertible so there's a IDFT

- symmetry of DFT: let x be a real signal and let $y = \text{DFT}(x)$

  $$\overline{y[k]} = y[N - k] \qquad\qquad 0 < k < N$$

- sinusoids: a sinusoids is a vector $x \in \mathbb{C}^N$ of the form

  $$x = A\cos(2\pi kt + \phi)$$
  $$A = \text{amplitude} \qquad k = \text{frequency} \qquad \phi = \text{phase}$$

  - let $t = (0, 1/N, 2/N, \ldots, (N-1)/N)$
  - for any sinusoids $x = A\cos(2\pi kt + \phi)$

  $$\text{DFT}(x) = \frac{AN}{2}(e^{i\phi} e_k + e^{-i\theta} e_{N-k})$$

  - **basically**: if your signal is a sin/cos wave then you will only have 2 non-0 entries in the frequency domain (and they're complex conj)
  - note: $e^{i\phi}$ is Euler's number, e_k is the standard basis (meaning we only care about certain entries)
  - equivalently, we can say

  $$A\cos(2\pi kt + \phi) = \frac{A}{2}e^{i\phi} f_k + \frac{A}{2}e^{-i\phi} f_{N-k}$$