Eigenvalues and Eigenvectors

$$A_{n\times n} x_i = \lambda_i x_i \quad , \quad i = 1, 2, \dots, n$$

$$[Ax_1, Ax_2, \dots, Ax_n] = [\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_n x_n]$$

$$A \underbrace{[x_1, \dots, x_n]}_{P} = \underbrace{[x_1, \dots, x_n]}_{P} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{D}$$

$$AP = PD \Rightarrow A = PDP^{-1} \quad \text{then} \quad A^k = PD^kP^{-1}$$

Recall: For a symmetrical matrix $A_{n\times n}$

- all eigenvalues of A are real
- eigenvectors belonging to distinct eigenvalues must be orthogonal

$P = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n]$ has orthonormal basis.

$P^TP = I \Rightarrow P^T = P^{-1}$ $\qquad P^{-1}AP = D \Rightarrow P^TAP = D \Rightarrow A = PDP^T$

If A is positive definite, $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}^2$

So $A = PDP^T = P\sqrt{D}\sqrt{D}P^T = (P\sqrt{D}P^T)(P\sqrt{D}P^T) = L^2$.

Inner Product

$(v,w) \mapsto \langle v, w\rangle$ s.t. it is bilinear, symmetric, positive definite

Let $v = w$. $\langle v, v\rangle = \|v\|^2 \Rightarrow \|v\| = \sqrt{\langle v, v\rangle}$

A vector space over a field $\mathbb{K}$ is sometimes called a <u>linear space</u>.

$\mathbb{K} = \mathbb{C}$ Notation: $V/\mathbb{C}$

$V = \mathbb{C}^n/\mathbb{C}$ $\quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ $\quad w = \begin{bmatrix} w_1 \\ w_n \end{bmatrix}$ $\quad \langle z, w \rangle = z_1\overline{w}_1 + \dots + z_n\overline{w}_n$

let $n \to \infty$

(i) $\langle z, z \rangle = z_1\overline{z}_1 + \dots + z_n\overline{z}_n = \|z_1\|^2 + \dots + \|z_n\|^2 \geq 0$

(ii) $\overline{\langle z, w \rangle} = \overline{z_1\overline{w}_1 + \dots + z_n\overline{w}_n} = \overline{z}_1 w_1 + \dots + \overline{z}_n w_n$

(iii) $\langle cz_1 + z_2, w \rangle = c\langle z_1, w \rangle + \langle z_2, w \rangle$

By (ii) and (iii), $\langle z, c\,w_1 + w_2 \rangle = \overline{c}\langle z, w_1 \rangle + \langle z, w_2 \rangle$

<u>Def:</u> Let X be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. A <u>scalar or inner product</u> on X is a mapping

$$\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$$

satisfying for any $k \in \mathbb{K}$, $x, x', y \in X$

(i) $\langle x, x \rangle \geq 0$ and "$=0$" iff $x = 0$

(ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$

(iii) $\langle kx + x', y \rangle = k\langle x, y \rangle + \langle x', y \rangle$

<u>Ex:</u> $\ell^2(\mathbb{R}) = \{\{x_i\}_{i=0}^{\infty}, x_i \in \mathbb{R}$ and $\sum x_i^2 < \infty\}$

$L^2[a,b] = \{f : f$ is a fcn $: [a,b] \to \mathbb{R}$ and $\int_a^b f^2(x)\,dx$ exists and is finite$\}$

$\langle f, g \rangle = \int_a^b fg\,dx$

$\ell^2(\mathbb{C}) = \{\{z_i\}_{i=0}^{\infty}, z_i \in \mathbb{C}, \sum |z_i| < \infty\}$

Define $\langle \{z_i\}_{i=1}^{\infty}, \{w_i\}_{i=1}^{\infty} \rangle = \sum_{i=0}^{\infty} z_i \overline{w_i}$

Recall: Totally ordered

a relation R ($\leq$) on a set A

1) antisymmetric $\forall a,b \in A$ if $a \leq b$ and $b \leq a \Rightarrow a = b$

2) transitive $\forall a,b,c \in A$ if $a \leq b$, $b \leq c \Rightarrow a \leq c$

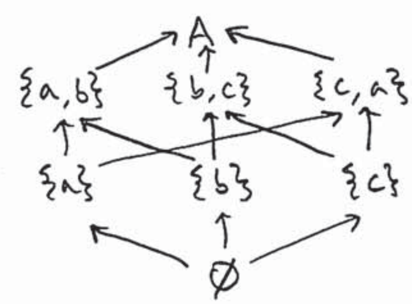
3) $a \leq b$ or $b \leq a$ (connex property)

Partially ordered

Ex: $A = \{a, b, c\}$ "$\subseteq$" is a partial order on $P(A)$ but not a total order

inclusion — power set

Hasse diagram



$$\underbrace{\emptyset \subseteq \{b\} \subseteq \{b, c\} \subseteq A}_{\text{called a chain}}$$

Can't compare $\{b\}$ and $\{c\}$ so only partially ordered.

Zorn's Lemma: Let P be a partially ordered set s.t. every chain has an upper bound in P. Then the set P contains at least one maximum element.

Proof idea: Use Hasse diagram for (P, R). Suppose there were no such maximum element. So $\exists$ chain in Hasse diagram where we can add larger and larger elements since nothing is maximum. But this would not be bounded above.

Recall: If we have an orthonormal basis of a vector space $V/\mathbb{R}$,

$$\vec{v} = a_1\vec{u}_1 + \dots + a_n\vec{u}_n \qquad a_i = \vec{v} \cdot \vec{u}_i, \quad i = 1, \dots, n$$

Linear Functional

<u>Def:</u> Let V be a vector space over the field $\mathbb{F}$. A linear mapping $l: V \to \mathbb{F}$ is also called a <u>linear functional</u> on V.

<u>Ex:</u> Define $l: \mathbb{F}^n \to \mathbb{F}$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \longmapsto a_1 x_1 + \dots + a_n x_n = l(x) \quad , \quad a_i \in \mathbb{F} \ , \ i = 1, \dots, n$$

<u>Ex:</u> $\operatorname{tr} A = a_{11} + \dots + a_{nn}$, $l: \mathbb{F}^{n \times n} \to \mathbb{F}$, $\operatorname{tr}(cA + B) = c\operatorname{tr}(A) + \operatorname{tr}(B)$

<u>Ex:</u> "Evaluation at a point" is a linear functional on $P_n(\mathbb{C})$

$$L_{t_0}: P_n(\mathbb{C}) \to \mathbb{C} \ , \ t_0 \in \mathbb{C} \qquad p(t) \mapsto L_{t_0}(p(t)) = p(t_0)$$

$$V/\mathbb{K} \xrightarrow{f} W/\mathbb{K} \qquad f \text{ is linear if } f(v+w) = f(v) + f(w) \text{ and } f(cv) = cf(v)$$

If V is finite, for $L: V \to V$, the <u>spectrum</u> is the set of eigenvalues.

<u>Def:</u> A complex number λ is said to be in the spectrum of a bounded linear operator T if $\lambda I - T$ is not invertible.

Spectral Mapping Theorem

<u>Thm:</u> μ is an eigenvalue of $q(A) \iff \exists \lambda$ (eigenvalue of A) s.t. $\mu = q(\lambda)$

Note: A and $q(A)$ share the same set of eigenvalues.

Cayley - Hamilton Theorem

<u>Thm:</u> Every matrix A satisfies its own characteristic equation.

$$p_A(A) = 0$$

Proof: $p_A(\lambda) = \lambda^n - (a_{11}+a_{22}+\dots+a_{nn})\lambda^{n-1} + \dots + (-1)^n \det A$

$$= \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$$

For $\lambda I - A$, consider the adjoint of A, denoted $B(\lambda)$. Since each entry of $B(\lambda)$ is a cofactor of $\lambda I - A$, the degree is $\leq n-1$. We write

$$B(\lambda) = \lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \dots + \lambda B_{n-2} + B_{n-1}$$

Note that the B_i's are scalar matrices. We have

$$(\text{adjoint } M)M = (\det M)I, \text{ where } M \text{ is } n\times n \text{ matrix}$$

In our case, $B(\lambda)(\lambda I - A) = \det(\lambda I - A)I$.

$$\text{RHS} = (\lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \dots + a_{n-1}\lambda + a_n)I$$

$$\text{LHS} = (\lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \lambda^{n-3}B_2 + \dots + \lambda^2 B_{n-2} + \lambda B_{n-1})$$

$$= \lambda^n B_0 + \lambda^{n-1}B_1 + \dots + \lambda^2 B_{n-2} + \lambda B_{n-1}$$
$$- \lambda^{n-1}B_0 A - \dots - \lambda^2 B_{n-3}A - \lambda B_{n-2}A$$

Comparing the left and right hand sides gives

$$\begin{aligned}
B_0 &= I \\
B_1 - B_0 A &= a_1 I \\
B_2 - B_1 A &= a_2 I \\
&\vdots \\
B_{n-2} - B_{n-3}A &= a_{n-2}I \\
B_{n-1} - B_{n-2}A &= a_{n-1}I \\
-B_{n-1}A &= a_n I
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
B_0 A^n &= A^n \\
B_1 A^{n-1} - B_0 A^n &= a_1 A^{n-1} \\
B_2 A^{n-2} - B_1 A^{n-1} &= a_2 A^{n-2} \\
&\vdots \\
B_{n-2}A^2 - B_{n-3}A^3 &= a_{n-2}A^2 \\
B_{n-1}A - B_{n-2}A^2 &= a_{n-1}A \\
-B_{n-1}A &= a_n I
\end{aligned}$$

Adding them up gives $p_A(A) = 0$.

Def: A matrix $U \in M$ is called <u>unitary</u> if $UU^* = I \ (= U^*U)$ where $U^* \triangleq (\bar{U})^T$. If U is a real matrix ($U^* = U^T$) then U is called an <u>orthogonal</u> matrix.

Prop: (i) U is invertible ($U^{-1} = U^*$)
(ii) $|\det U| = 1$

Thm: U is unitary $\Leftrightarrow \|Ux\| = \|x\| \Leftrightarrow \exists$ orthonormal system $\{u_1, u_2, \dots, u_n\}$ s.t. $\langle u_i, u_j \rangle = \delta_{ij}$

Schur's Theorem

Given $A \in M_n(\mathbb{C})$ with eigenvalues $\lambda_1, \dots, \lambda_n$ (could be complex), counting multiplicities, then $\exists$ unitary matrix $U \in M_n(\mathbb{C})$ s.t.

$$A = U \underbrace{\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\text{upper triangular}} U^*$$

Proof: By induction, $n=1$ is true. Suppose for $n-1 \times n-1$ matrix $\tilde{A}$, $\exists \tilde{W}$ (unitary) s.t.

$$\tilde{W}^* \tilde{A} \tilde{W} = \begin{bmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Now for $n \times n$ matrix A, let $\lambda_1, \lambda_2, \dots, \lambda_n$ be eigenvalues. Extend $v_1 \neq 0$ into orthonormal basis $\{v_1, v_2, \dots, v_n\}$.

$$A v_1 = \lambda_1 v_1$$
$$A v_2 = b_{11} v_1 + b_{12} v_2 + \dots + b_{1n} v_n$$
$$\vdots$$

$$(Av_1, \dots, Av_n) = (v_1, \dots, v_n) \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \tilde{A} & \\ 0 & & & \end{array}\right]$$

$$A = V \left[\begin{array}{c|c} \lambda_1 & * \cdots * \\ \hline 0 & \tilde{A} \end{array}\right] V^*$$

Let $W = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{W} \end{array}\right]$. Then we see that $W^*W = I_n$ and

$$W^*AW = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Thm: Given $A \in M_n$ and $\varepsilon > 0$, there $\exists$ diagonalizable matrix $\tilde{A} \in M_n(\mathbb{C})$ s.t.

$$\sum_{1 \le i,j \le n} |a_{ij} - \tilde{a}_{ij}|^2 < \varepsilon$$

Dual Spaces

Consider the set of all linear functionals $l : W \to \mathbb{F}$ denoted $L(W, F)$. This is called the dual space of W, denoted W^*.

Thm: $\dim W^* = \dim W$

Proof: Method 1

$L(W,V) \cong \{A_{m \times n}\} \cong M_{\mathbb{F}}(n,m)$. In our case $M_{\mathbb{F}}(n,1)$, $\dim M_{\mathbb{F}}(n,1) = n$.

Method 2

Recall $\{u_1, \dots, u_n\}$ orthonormal $v_1 = a_1 u_1 + \dots + a_n u_n \Rightarrow v_1 \cdot u_1 = a_1$. We want to construct a basis for W^* which contains n basis vectors $\{l_1, \dots, l_n\}$

$$\begin{array}{llll} l_1(v_1) = 1 & l_2(v_1) = 0 & & l_n(v_1) = 0 \\ l_1(v_2) = 0 & l_2(v_2) = 1 & \dots & l_n(v_2) = 0 \\ \vdots & \vdots & & \vdots \\ l_1(v_n) = 0 & l_2(v_n) = 0 & & l_n(v_n) = 1 \end{array}$$

$l_i(v_j) = \delta_{ij}$ where $\{v_1, \dots, v_n\}$ is a basis of W

We form $a_1 l_1 + \dots + a_n l_n = 0$ for contradiction.

$$(a_1 l_1 + \dots + a_n l_n)(v_i) = 0(v_i)$$

$$\Rightarrow a_1 l_1(v_i) + a_2 l_2(v_i) + \dots + a_n l_n(v_i) = 0$$

$$\Rightarrow a_i = 0, \ i = 1, 2, \dots, n$$

From method 1, $\dim W^* = n$ so $\{l_1, \dots, l_n\}$ form basis of W^*

Consider the dual space of V^* (dual space of W). What is V^{**} (or V'')?

$V'' = L(V', \mathbb{K})$. Thm: $V'' \cong V$ (natural isomorphism)

Proof: $V \to V''$ $\quad \varphi : x \mapsto L_x \quad \begin{bmatrix} L_x : V' \to \mathbb{K} \\ L_x(l) = l(x) \end{bmatrix}$

Wts (i) φ is bijection

(ii) φ preserves linear structure (i.e. $\varphi(cx+y) = c\varphi(x) + \varphi(y)$)

<u>Thm:</u> Given $A \in M_n$ and $\varepsilon > 0$, there exists a diagonalizable matrix $\tilde{A} \in M_n$ s.t.

$$\sum_{1 \le i,j \le n} |a_{ij} - \tilde{a}_{ij}|^2 < \varepsilon$$

<u>Proof:</u> Use Schur's Thm. We have $\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$. Then,

$$A = U \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} U^* \qquad \text{Key: } \tilde{\lambda}_i = \lambda_i + i\eta, \ i = 1, \dots, n$$

We then form $\tilde{A} = U \begin{bmatrix} \tilde{\lambda}_1 & * \\ 0 & \tilde{\lambda}_n \end{bmatrix} U^*$. Then,

$$tr((A - \tilde{A})^*(A - \tilde{A})) = tr\left[U \begin{bmatrix} \overline{\lambda_1 - \tilde{\lambda}_1} & & \\ & \ddots & \\ & & \overline{\lambda_n - \tilde{\lambda}_n} \end{bmatrix} U^* U \begin{bmatrix} \lambda_1 - \tilde{\lambda}_1 & & \\ & \ddots & \\ & & \lambda_n - \tilde{\lambda}_n \end{bmatrix} U^* \right]$$

$$= \sum |\lambda_i - \tilde{\lambda}_i|^2 = \sum i^2 \eta^2 < \varepsilon$$

let $\eta^2 = \frac{\varepsilon}{\sum i^2}$

Euclidean Structure (Inner Product Space)

$(V/\mathbb{F}, \langle , \rangle) \qquad \langle , \rangle : V \times V \to \mathbb{F}$

1) linear on the first coordinate
2) conjugate symmetry $\langle x, y \rangle = \overline{\langle y, x \rangle}$
3) $\langle v, v \rangle \ge 0$ and "$= 0$" holds iff $v = 0$

Two key facts:

① Cauchy-Schwarz inequality

② $\exists$ an orthonormal basis

Hilbert Space

<u>Def:</u> An inner product space X is called a <u>Hilbert space</u> if every Cauchy sequence converges to a vector in X.

Let $\mathcal{L}(\mathbb{X}) = \{T : \mathbb{X} \to \mathbb{X} \text{ linear map}\}$. $\dim \mathbb{X} = n$

Q: Can we give a norm to T?

Let $\{x_1, \dots, x_n\}$ be an orthonormal basis. Then $\forall x \in \mathbb{X}$,

$x = \sum a_i x_i$.

$$\|Tx\| = \|\sum a_i T x_i\| \le \sum |a_i| \|Tx_i\| \le M \sum |a_i|$$

where $M = \max\{\|Tx_1\|, \|Tx_2\|, \dots, \|Tx_n\|\}$. Then,

$$M \sum |a_i| \le M \sum \sqrt{|a_i|^2} \sqrt{n} = M \|x\| \sqrt{n}$$

So, $\frac{\|Tx\|}{\|x\|} \le M$. Let $\vec{u} = \frac{\vec{x}}{\|\vec{x}\|} \in S^{n-1}$ (which is compact). Then

$\|T\vec{u}\| \le M$ so $\exists$ max on compact S^2

<u>Def:</u> $\|T\| = \max\{\|T\vec{u}\|, \|\vec{u}\| = 1\}$. In general, $\|T\| = \sup_{\vec{u}}\{\|T\vec{u}\|, \|\vec{u}\| = 1\}$

Riesz Representation Theorem

Q: What does $\mathbb{X}^*$ look like if we put Euclidean structure on $\mathbb{X}/\mathbb{F}$?

Let $\{x_1, \dots, x_n\}$ be an orthonormal basis of $\mathbb{X}$. Then $\forall x$,

$x = a_1 x_1 + \dots + a_n x_n$ and $a_i = \langle x, x_i \rangle$. Let $\ell \in \mathbb{X}^*$. If $x = k_1 x_1 + \dots + k_n x_n$,

$$\ell(x) = k_1 \ell(x_1) + \dots + k_n \ell(x_n) = a_1 \langle x, x_1 \rangle + a_2 \langle x, x_2 \rangle + \dots + a_n \langle x, x_n \rangle$$

$$= \langle x, \bar{a}_1 x_1 + \bar{a}_2 x_2 + \dots + \bar{a}_n x_n \rangle = \langle x, y \rangle \text{ where } y = \bar{a}_1 x_1 + \dots + \bar{a}_n x_n.$$

<u>Thm:</u> (Riesz Representation Theorem) For $\ell \in \mathbb{X}'$, $\exists y \in \mathbb{X}$ s.t. $\ell(x) = \langle x, y \rangle$

Self-adjoint Map

<u>Def:</u> $T : \mathbb{X} \to \mathbb{X}$ is called <u>self-adjoint</u> if $T^* = T$, i.e.

$\langle x, Tx \rangle = \langle Tx, x \rangle$, $\forall x \in \mathbb{X}$.

Let M be a matrix representation of T with respect to some basis B. For complex $n \times n$ matrix M, M is self-adjoint if $M^* = M$ (or Hermitian) (i.e. $\overline{M^T} = M$)

Adjoint of a Linear Map

Let $\mathbb{X}$ be an inner product space. Consider $T: \mathbb{X} \to \mathbb{X}$. The dual is $T': \mathbb{X}' \to \mathbb{X}'$ where $\ell \mapsto \ell \circ T$.

$$\begin{tikzcd} \mathbb{X} \arrow[r, "T"] \arrow[rr, bend right, "\ell \circ T"'] & \mathbb{X} \arrow[r, "\ell"] & \mathbb{R} \end{tikzcd}$$

$$\begin{tikzcd} \ell \in \mathbb{X}' \arrow[r, "T'"] \arrow[d, leftrightarrow] \arrow[r, bend left, "T'(\ell) = \ell \circ T"] & \mathbb{X}' \\ y \in \mathbb{X} \arrow[r, "T^*"'] \arrow[u, phantom, "\cong"] & \mathbb{X} \arrow[u, phantom, "\cong"] \end{tikzcd}$$

$T^* y \triangleq z_y$

$\mathbb{X}' \cong \mathbb{X}$

$\ell \leftrightarrow y$

$T'(\ell)$

$T^*: \mathbb{X} \to \mathbb{X}$

$y \mapsto z_y$

We claim T^* is a linear map from $\mathbb{X}$ to $\mathbb{X}$ and $\langle Tx, y \rangle = \langle x, T^* y \rangle$, $\forall x, y \in \mathbb{X}$. T^* is called the <u>adjoint</u> of T.

To get a more concrete idea of the adjoint, let's take a look at it's matrix representation.

Let $\{x_1, \dots, x_n\}$ be an O.N. basis of $\mathbb{X}$

$$T x_j = \sum_{i=1}^{n} m_{ij} x_i \quad T \leftrightarrow (m_{ij}) = M$$

$$T^* x_j = \sum_{i=1}^{n} n_{ij} x_i \quad T^* \leftrightarrow (n_{ij}) = N$$

How are M and N related?

$$\langle T x_j, x_i \rangle = m_{ij} = \langle x_j, T^* x_i \rangle$$
$$\langle T^* x_j, x_i \rangle = \overline{\langle x_i, T x_j \rangle} = \overline{m_{ji}}$$

$$\Rightarrow n_{ij} = \overline{m_{ji}} \Rightarrow N = \overline{(M^T)} = M^*$$

<u>Def</u>: M^* is called the <u>conjugate transpose</u> of M.

Thm: Let X be an inner product space and T be self adjoint (i.e. $T^* = T$ or $\langle Tx, y\rangle = \langle x, Ty\rangle$). Then T has real eigenvalues and a set of eigenvectors that form an orthonormal basis for X.

Proof: Suppose $a+bi$ is an eigenvalue of T so $\exists x \neq 0$ s.t. $Tx = (a+bi)x$. Then, $(T-aI)x = ibx$. Now, $(T-aI)^* = T^* - aI^*$ $= T - aI$ so $T-aI$ is also self adjoint.

$$\begin{aligned} \langle (T-aI)x, x\rangle &= \langle x, (T-aI)x\rangle \\ = \langle ibx, x\rangle \quad &\quad = \langle x, ibx\rangle \\ = ib\langle x, x\rangle \quad &\quad = \overline{ib}\langle x, x\rangle \\ = ib\|x\|^2 \quad &\quad = -ib\|x\|^2 \end{aligned} \quad \Rightarrow \quad \begin{aligned} &2ib\|x\|^2 = 0 \\ &\Rightarrow 2ib = 0 \quad (x \neq 0) \\ &\Rightarrow b = 0 \end{aligned}$$

Now show "eigenvectors belonging to distinct eigenvalues are orthogonal". Say $Tx = a_i x$, $a_i \in \mathbb{R}$, $x \neq 0$ and $Ty = a_j y$, $a_j \in \mathbb{R}$, $y \neq 0$.

$$\begin{aligned} \langle Tx, y\rangle &= \langle a_i x, y\rangle = a_i \langle x, y\rangle \\ &\shortparallel \\ \langle x, Ty\rangle &= \langle x, a_j y\rangle = a_j \langle x, y\rangle \end{aligned} \quad \Rightarrow \quad \begin{aligned} &\underbrace{(a_j - a_i)}_{\neq 0 \text{ since } a_i \neq a_j}\langle x, y\rangle = 0 \\ &\Rightarrow \langle x, y\rangle = 0 \end{aligned}$$

Thm: Let M be a real self-adjoint matrix, then $\exists$ O.N. matrix P s.t

$$P^* M P = D$$

Spectral Resolution of a Self Adjoint Map

Let $T: X \to X$ be self adjoint. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues. We have $X = N_1 \oplus \ldots \oplus N_k$. $\forall x \in X$, $x = x^{(1)} + \ldots + x^{(k)}$, $x^{(i)} \in N_i$.

① For $i = 1, \ldots, k$, define projection $P_i : X \to N_i$, with identity $= \sum_{i=1}^{k} P_i$. (call it the "identity resolution")

② $P_i^2(x) = P_i(x^{(i)}) = x^{(i)} = P_i(x) \implies P_i^2 = P_i$

③ For $i \neq j$, $P_i P_j(x) = P_i(x^{(j)}) = 0$, $\forall x \Rightarrow P_i P_j = 0$

④ $\langle x, P_j(y)\rangle = \langle \sum_{i=1}^{k} P_i(x), P_j(y)\rangle = \sum_{i=1}^{k} \langle P_i(x), P_j(y)\rangle = \langle P_j(x), P_i(x)\rangle$

If we do the same for $\langle P_j(x), y\rangle$, we see that it is self adjoint.

$$Tx = T\left(\sum_{i=1}^{k} P_i(x)\right) = \sum_{i=1}^{k} T(P_i x) = \sum_{i=1}^{k} Tx^{(i)} = \sum_{i=1}^{k} \lambda x^{(i)} = \sum_{i=1}^{k} \lambda_i P_i(x), \ \forall x \in \mathbb{X}.$$

$T = \sum_{i=1}^{k} \lambda_i P_i$ is called the <u>spectral resolution</u> of T. We also have $e^{sT} = \sum_{i=1}^{k} e^{\lambda_i s} P_i$ where $e^A = I + A + \frac{A^2}{2!} + \dots$. We also have $T^n = \sum_{i=1}^{k} \lambda_i^n P_i$.

Normed Linear Vector Space

<u>Def:</u> A linear (or vector) space $\mathbb{X}$ over $\mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$) is a <u>normed linear space</u> if $\exists \ \|\cdot\|$ (or $|\cdot|$) $: \mathbb{X} \to \mathbb{R}$ satisfying

1) for $x \in \mathbb{X}$, $|x| \geq 0$ and "$= 0$" iff $x = 0$
2) $|x+y| \leq |x| + |y|$
3) $|kx| = |k||x|$, $\forall k \in \mathbb{K}$, $x \in \mathbb{X}$

<u>Ex:</u> $\mathbb{X} = \mathbb{R}^n$, $\|x\| = \sqrt{\langle x, x\rangle}$

$\mathbb{X} = \mathbb{K}^n$, $|x| = |(x_1, \dots, x_n)|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$

$\mathbb{X} = \mathbb{K}^n$, $|x|_1 = \sum_{i=1}^{n} |x_i|$

$\mathbb{X} = \mathbb{K}^n$, $|x|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$

Hölder's Inequality

<u>Thm:</u> $|x \cdot y| \leq |x|_p |y|_q$, where $1/p + 1/q = 1$

This reduces to the Cauchy-Schwarz inequality when $p = q = 2$.

(Q:) How are all these different norms related?

Def: Norms $|\cdot|$ and $\|\cdot\|$ on $\mathbb{X}$ are called <u>equivalent</u> if there exists a constant c s.t. $\forall x \in \mathbb{X}$, $|x| \le c\|x\|$ and $\|x\| \le c|x|$.

Thm: Any two norms on a finite dimensional linear vector space $\mathbb{X}$ are equivalent.

Banach Space

Def: A <u>Banach space</u> is a vector space over a field $\mathbb{K}$ (say $\mathbb{K} = \mathbb{R}^n$ or $\mathbb{C}^n$) which is equipped with a norm $\|\cdot\|_{\mathbb{X}}$ and which is complete with respect to $\|\cdot\|_{\mathbb{X}}$.

Ex: Consider the space of functions $C[a,b]$ = set of all continuous functions on $[a,b]$. Let $\|f\| = \sup_{x \in \mathbb{R}} |f(x)|$. This norm is not from an inner product.

Every Hilbert space is a Banach space but the converse is not true.

Bounded Linear Operator

Def: A bounded linear operator (or transform) from $(V, |\cdot|)$ to $(W, \|\cdot\|)$ if there exists $M > 0$ s.t. $\|L(v)\|_w \le M|v|_v$ or $\|L(v)\|_w / |v|_v \le M$.

Ex: The shift operator on the ℓ^2 space of all sequences $\{x_n\}_n$.

$$L(x_0, x_1, x_2, \dots) = (0, x_0, x_1, \dots)$$

is bounded. Its operator norm $= 1$.

$K : [a,b] \times [c,d] \longrightarrow \mathbb{R}$ $\quad L(f(y)) = \int_a^b K(x,y) f(x)\, dx$

$\Rightarrow L$ is bounded.

Schatten p-norm

Def: Let H_1, H_2 be separable Hilbert spaces, and T be a linear bounded operator from H_1 to H_2 for $p \in [1, \infty)$. Define the Schatten p-norm of T as

$$\|T\|_p = \left(\sum_n s_n^p(T) \right)^{1/p}$$

for $s_1(T) \geq s_2(T) \geq \ldots \geq s_n(T) \geq \ldots \geq 0$ the singular values of T.

Note for $p=2$, $\|T\|_2 = \sum_n (\sqrt{\lambda_i})^2 = \lambda_1 + \ldots + \lambda_n = \operatorname{tr} T$. In general, $\|T\|_p^p = \operatorname{tr}(|T|^p)$

Fréchet Derivative

Def: Let $U \subseteq (V, \|\cdot\|_v)$ and $(W, \|\cdot\|_w)$ be normed spaces. Then $f: U \to W$ is called Fréchet differentiable at $x \in U$ if there exists a bounded linear operator $A: V \to W$ s.t.

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Ah\|_w}{\|h\|_v} = 0$$

Recall for $f: U \subseteq \mathbb{R}^n \to \mathbb{R}$, $f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2!}(x-x_0)^T \nabla^2 f(x_0)(x-x_0) + \ldots$

Here, $f(x+h) = f(x) + Ah + o(h)$.

If there exists such an operator A, it is unique $Df(x) = A$ and call it the Fréchet derivative of f at x. Say f is C^1 if $Df: U \to B(V,W)$, $x \mapsto Df(x): V \to W$ continuous of each value of x

Thm: f is $C^1 \Rightarrow f$ is differentiable

Matrix Calculus

Given dimension matrix-valued functions of matrix variable $f(x)$ and $g(x)$

$$\nabla_x \left[f(x)^T g(x) \right] = \nabla_x(f)\, g + \nabla_x(g)\, f$$

The set of all probability distributions forms a certain space (locally looks like a vector space)

Hellinger Distance

Let P and Q be two probability distributions and let p and q be their density functions

$$H^2(P,Q) = \frac{1}{2}\int(\sqrt{p}-\sqrt{q})^2 dx = 1 - \int\sqrt{p(x)q(x)}\,dx$$

For discrete distributions,

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}(\sqrt{p_i}-\sqrt{q_i})^2}$$

where $P=(p_1,\dots,p_k)$, $Q=(q_1,\dots,q_k)$ are probability distributions.

KL Divergence

$$D_{KL}(P\|Q) = -\sum_i P(i)\log\left(\frac{Q(i)}{P(i)}\right)$$

Bhattacharyya Distance

$$D_B(p,q) = -\ln(BC(p,q)) \quad \text{where} \quad BC(p,q) = \sum_{x\in X}\sqrt{p(x)q(x)}$$

Raleigh Quotient

If M is positive definite, we can use M to define an inner product

$$\langle x, Mx\rangle = x^T Mx$$

<u>Def</u>: The Raleigh quotient R of M denoted by $R_M: x\setminus\{0\}\to\mathbb{R}$ by

$$R(x) = R_M(x) = \frac{q(x)}{p(x)} = \frac{(x,Mx)}{(x,x)} \quad \text{for} \quad x\neq 0.$$

The key is we can use $R(x)$ to calculate eigenvalues if we estimate eigenvectors.

Positive Matrices and Applications to Markov Processes

<u>Def</u>: A matrix $P=(p_{ij})_{n\times n}$ is <u>positive</u> if $p_{i,j}>0$, $\forall i,j$.

① For $x=[x_1\ x_2\ \dots\ x_n]^T$, $y=[y_1\ y_2\ \cdots\ y_n]^T$, we say $x<y$ iff $x_i<y_i$, $\forall i$.

② We say $x\le y$ iff $x_i\le y_i$, $\forall i$.

Note that $x \leq y \not\Leftrightarrow x < y$ or $x = y$.

③ If $x \geq 0$, we say x is a non-negative vector.

④ Let $\xi_0 = (1, 1, \ldots, 1) \in \mathbb{R}$ and $x \geq 0$. We say x is L_1-normalized if $\xi_0 x = (1\ 1 \ldots 1)(x_1 \ldots x_m)^T = \sum_{i=1}^{m} x_i = 1$.

Perron's Theorem

Thm: If P is a positive matrix, then P has a dominant eigenvalue $\lambda(P)$ s.t.

1) $\lambda(P) > 0$, $\lambda(P)$ is an eigenvalue of P and $\exists h > 0$ s.t. $Ph = \lambda(P)h$.
2) $\lambda(P)$ is simple
3) For any eigenvalue μ, $|\mu| \leq \lambda(P)$
4) For any eigenvalue $\mu \neq \lambda(P)$, $|\mu| < \lambda(P)$ and if (μ, f) is an EV pair, then $f \not> 0$.

Frobenius Theorem

Thm: Let $P \geq 0$ be an $n \times n$ matrix. Then there exists $\lambda(P) \in \mathbb{R}$ s.t.

1) $\lambda(P)$ is an eigenvalue of P, $\lambda(P) \geq 0$ and $\exists h \geq 0$ s.t. $\xi_0 h = 1$ and $Ph = \lambda(P)h$.
2) If $k \in \mathbb{C}$ is an eigenvalue of P
   a) $|k| \leq \lambda(P)$
   b) $|k| = \lambda(P) \Rightarrow k = e^{\frac{2\pi i k}{m}} \lambda(P)$

Applications to Advanced ML

Concentration inequalities deal with the derivation of a function of independent random variables from their expectation.

The law of large numbers from probability theory states that the sum of the independent random variables are, under very mild conditions, close to their expectation with a large probability.

Markov Inequality

If $Z \geq 0$, then $\mathbb{P}\{Z \geq \mathbb{E}Z + t\} \leq \frac{\mathbb{E}Z}{t}$

Chebyshev's Inequality

$\mathbb{P}(|x-\mu| \geq a) \leq \frac{Var(x)}{a^2}$

Applications of Chebyshev's Inequality

(I) Weak Law of Large Numbers (WLLN)

$$\lim \mathbb{P}(|\bar{x}_N - \mu| > \varepsilon) = 0$$

sample mean ↗ ↖ population mean

Proof: Use Chebyshev's inequality.

$$\mathbb{P}(|\bar{x}_N - \mu| > \varepsilon) \leq \frac{Var(\bar{x}_N)}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2} \xrightarrow[N\to\infty]{} 0$$

$$\bar{X}_N = \frac{1}{N}\sum x_i \qquad Var(\bar{X}_N) = \frac{1}{N^2}\sum Var(x_i) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

(II) Chernoff Bounds

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{t\cdot x} \geq e^{t\cdot a}) \leq \frac{\mathbb{E}[e^{t\cdot x}]}{e^{t\cdot a}}$$

Bounded Difference Inequalities

Let $x_1, \dots, x_n$ be independent random variables. Suppose

$$\sup_{x_1,\dots,x_n,x_i'} |g(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) - g(x_1,\dots,x_i',\dots,x_n)| \leq c_i$$

for $i = 1, 2, \dots, n$. Then

$$\mathbb{P}\Big(g(x_1,\dots,x_n) - \mathbb{E}(g(x_1,\dots,x_n)) \geq \varepsilon\Big) \leq \exp\left\{-\frac{2\varepsilon^2}{\sum_{i=1}^{n} c_i^2}\right\}$$