8. Self-adjoint mappings:

Throughout, let X be a finite-dimensional Euclidean space.

Def: Recall that a linear map $M: X \to X$ is <u>self-adjoint</u> (or <u>Hermitian</u>) if $M^* = M$. It is <u>anti-self-adjoint</u> (or <u>anti-Hermitian</u>) if $M^* = -M$.

Remark: Every linear map $M: X \to X$ can be decomposed into a self-adjoint part and an anti-self-adjoint part, by

$$M = H + A, \qquad H = \frac{M+M^*}{2}, \qquad A = \frac{M-M^*}{2}.$$

Indeed,
$$\begin{aligned} Re(x, Mx) &= \tfrac{1}{2}\left[(x,Mx) + \overline{(x,Mx)}\right] = \tfrac{1}{2}\left[(x,Mx) + (Mx,x)\right] \\ &= \tfrac{1}{2}\left[(x,Mx) + (x,M^*x)\right] = (x, Hx) \end{aligned}$$

$$\begin{aligned} Im(x, Mx) &= \tfrac{1}{2}\left[(x,Mx) - \overline{(x,Mx)}\right] = \tfrac{1}{2}\left[(x,Mx) - (Mx,x)\right] \\ &= \tfrac{1}{2}\left[(x,Mx) - (x,M^*x)\right] = (x, Ax). \end{aligned}$$

Quadratic forms

Motivation: Let $f(x_1, \ldots, x_n)$ be a real-valued function, $\mathbb{R}^n \to \mathbb{R}$. Recall the the <u>Taylor approximation</u> of f at $a \in \mathbb{R}^n$ up to 2^{nd} order says that for $y \in \mathbb{R}^n$ with $\|y\| \approx 0$,

$$f(a+y) \approx f(a) + \ell(y) + \tfrac{1}{2} q(y), \qquad \text{where}$$

* $f(a)$ is the 0^{th} order term
* $l(y)$ is the 1^{st} order term: $l(y) = (y, g)$ for some $g \in \mathbb{R}^n$

It turns out that $g = \nabla F = \begin{pmatrix} \partial F/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}$, the <u>gradient</u> of f.

* $q(y)$ is the 2^{nd} order term: $q(y) = \sum_{j=1}^{n} \sum_{i=1}^{n} h_{ij} y_i y_j$, where

$H = (h_{ij}) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i} \right)$ is the <u>Hessian</u> of f.

Note that H is self-adjoint, because $\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$,

and that $q(y) = [y_1, \ldots, y_n] \, H \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = (y, Hy)$.

Suppose $a \in \mathbb{R}^n$ is a critical point of f, i.e., $\nabla f = g = 0$.

Then the behavior of f is governed by the 2^{nd} order term $q(y)$.

<u>Def</u>: A function $q : X \to K$ of the form $q(x) = (x, Hx)$ for a self-adjoint map H is called a <u>quadratic form</u>.

Observe that $q(x) = [x_1, \ldots, x_n] \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{j=1}^{n} \sum_{i=1}^{n} h_{ij} x_i x_j$.

Suppose now that we can diagonalize H, that is, write

$H = P^{-1} D P$. Recall that this would mean that $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

and $P = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$, the matrix of eigenvectors of H.

Then, we would have

$$q(x) = (x, Hx) = x^T H x = x^T P^{-1} D P x.$$

Moreover, if P is real-valued and orthogonal, then $P^T P = I$, i.e, $P^{-1} = P^T$. Then we could put $z = Px$ and write

$$q(z) = z^T D z = [z_1, \ldots, z_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \sum_{i=1}^{n} \lambda_i z_i^2 .$$

This is much easier! Note that we can do this iff $P^T P = I$, i.e, iff X has an orthonormal basis of real eigenvectors of H. It turns out that this is always the case.

<u>Theorem 8.1</u>: A self-adjoint mapping $H: X \to X$ of a complex Euclidean space has only real eigenvalues, and a set of eigenvectors that forms an orthonormal basis of X.

<u>Proof</u>: It suffices to show that

(i) H has only real eigenvalues

(ii) H has no generalized eigenvectors (only genuine ones)

(iii) Eigenvectors corresponding to different eigenvalues are orthogonal.

<u>Pf</u>: (i) Let λ be an eigenvalue of H with eigenvector $v \neq 0$.

Then $(Hv, v) = (\lambda v, v) = \lambda (v, v)$

and $(v, Hv) = (v, \lambda v) = \bar{\lambda} (v, v)$

Since $(v,v) \neq 0$, $\lambda = \bar{\lambda} \Rightarrow \lambda$ is real. ✓

(ii) Suppose $(H - \lambda I)^d v = 0$. We must show $(H - \lambda I)v = 0$.

Induct on d. Base case $(d=2)$:

If $(H-\lambda I)^2 v = 0$, then $((H-\lambda I)^2 v, v) = 0$

$\Rightarrow ((H-\lambda I)v, (H-\lambda I)v) = \|(H-\lambda I)v\|^2 = 0 \Rightarrow (H-\lambda I)v = 0.$ ✓

Now, suppose $(H-\lambda I)^d v = 0 \Rightarrow (H-\lambda I)^2 \underbrace{(H-\lambda I)^{d-2} v}_{\text{call this } w} = 0$

We have $(H-\lambda I)^2 w = 0 \Rightarrow (H-\lambda I)w = 0$

$\Rightarrow (H-\lambda I)^{d-1} w = 0$

$\Rightarrow (H-\lambda I)v = 0$ (induction hypothesis) ✓

(iii) Suppose $Hv = \lambda v$, $Hw = \mu w$.

Then $\lambda(v,w) = (\lambda v, w) = (Hv, w) = (v, Hw) = (v, \mu w) = \mu(v,w)$

So if $\lambda \neq \mu$, then $(v,w) = 0$. ✓

□

<u>Corollary 8.2</u>: If H is self-adjoint, then $H = MDM^*$ for a diagonal matrix D and an <u>orthogonal</u> matrix M (that is, $M^*M = I$).

By Theorem 8.1, we can write $X = N^{(1)} \oplus \dots \oplus N^{(k)}$, where $N^{(i)}$ consists of eigenvectors with eigenvalue λ_i, and $\lambda_i \neq \lambda_j$ $(i \neq j)$.

Thus, we can write $x \in X$ as $x = x^{(1)} + \dots + x^{(k)}$, $x^{(i)} \in N^{(i)}$.

Note that $Hx = \lambda_1 x^{(1)} + \dots + \lambda_k x^{(k)}$.

Let $P_i(x)$ be the <u>projection</u> of x onto the eigenspace $N^{(i)}$, that is

$$P_i : X \longrightarrow X, \quad P_i : x \longmapsto x^{(i)}.$$

<u>Remark:</u> (a) $P_i P_j = 0$ if $i \neq j$ and $P_i^2 = P_i$

(b) $P_i^* = P_i$ (property of orthogonal projections).

<u>Def:</u> The decomposition $I = \sum_{i=1}^{k} P_i$ is called a <u>resolution of the identity</u>, and $H = \sum_{i=1}^{k} \lambda_i P_i$ is called the <u>spectral resolution of H</u>.

Corollary 8.2 can now be stated as follows:

<u>Theorem 8.3:</u> Let X be a complex Euclidean space, $H : X \longrightarrow X$ a self-adjoint linear map. Then there is a resolution of the identity and a spectral resolution of H.

It is now easy to define functions on H. For example,

$$H^2 = \sum_{i=1}^{k} \lambda_i^2 P_i, \quad H^m = \sum_{i=1}^{k} \lambda_i^m P_i,$$ and for any polynomial $p(t)$, we have $p(H) = \sum_{i=1}^{k} p(\lambda_i) P_i$.

Motivated by this, if f is any real-valued function defined on the spectrum (set of eigenvalues) of H, then we define

$$f(H) = \sum_{i=1}^{k} f(\lambda_i) P_i.$$

<u>Example</u>: $e^{H} = \sum_{k=1}^{k} e^{\lambda_i} P_i$.

<u>Theorem 8.4</u>: Suppose H and K are self-adjoint commuting maps.

Then they have a common spectral resolution, that is, there are orthogonal projections (as above) so that $I = \sum_{i=1}^{k} P_i$ and $H = \sum_{i=1}^{k} \lambda_i P_i$ and $K = \sum_{i=1}^{k} \mu_i P_i$.

<u>Proof</u>: Write $X = N^{(1)} \oplus \cdots \oplus N^{(k)}$, a product of eigenspaces of H corresponding to distinct eigenvectors.

Pick $N = N^{(j)}$. Then for every $x \in N$, $Hx = \lambda x$

$$\Rightarrow H(Kx) = K(Hx) = K\lambda x = \lambda(Kx)$$

Thus, Kx is an eigenvector of H, so K maps $N \longrightarrow N$.

Find a spectral resolution of K over N, i.e., write

$$K|_N = \sum_{i=1}^{k_j} \mu_{ji} P_{ji} \quad \text{and} \quad I|_N = \sum_{i=1}^{k_j} P_{ji}. \qquad \text{Assume } \mu_i\text{'s distinct.}$$

Note that $H|_N = \sum_{i=1}^{k_j} \lambda_{ji} P_{ji}$ (and $\lambda_{ji} = \lambda_j$ for each i).

Now, $N^{(j)} = N^{(j1)} \oplus N^{(j2)} \oplus \cdots \oplus N^{(jk_j)}$,

orthogonal eigenspaces of $K|_N$ (<u>and</u> of $H|_N$!)

Expanding each $N^{(j)}$ into eigenspaces of $K|_N$ gives a common spectral resolution of H and K, which we seek.

That is,

$$X = N^{(1)} \oplus N^{(2)} \oplus \dots \oplus N^{(k)}$$

$$\left(N^{(11)} \oplus \dots \oplus N^{(1k_1)}\right) \oplus \left(N^{(21)} \oplus \dots \oplus N^{(2k_2)}\right) \oplus \dots \oplus \left(N^{(k1)} \oplus \dots \oplus N^{(k\,k_k)}\right)$$

Note that not all of the corresponding eigenvalues will be distinct (and that's fine). $\square$

Remarks:

- This is easily generalized for any number of commuting maps.
- $(iM)^* = -iM^*$ (where $i = \sqrt{-1}$)

  Thus, if M is self-adjoint, then iM is anti-self-adjoint, and vice-versa. We can now conclude the following:

Corollary 8.5: Let A be an anti-self-adjoint mapping of a complex Euclidean space. Then

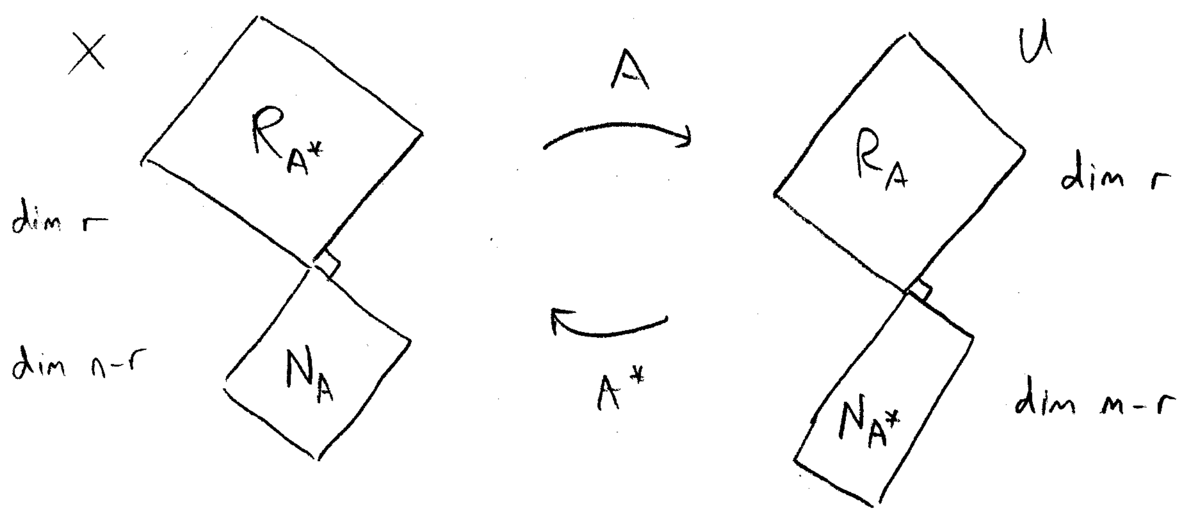
(a) The eigenvalues of A are purely imaginary

(b) X has an orthonormal basis of eigenvectors of A.

<u>Def</u>: A mapping $N : X \to X$ of a complex Euclidean space is <u>normal</u> if $NN^* = N^*N$.

<u>Remark</u>: Self-adjoint ($H^* = H$), anti-self-adjoint ($A^* = -A$), and unitary ($U^* = U^{-1}$) maps are all clearly normal.

Picture of this: Let $A : X \to U$ be linear.



<u>Facts</u> (proofs are HW):

- A restricted to R_{A^*} is a bijection $R_{A^*} \to R_A$
- $R_{A^*}^\perp = N_A$ and $R_A^\perp = N_{A^*}$

(and so $X = R_{A^*} \oplus N_A$ and $U = R_A \oplus N_{A^*}$.)

Think of R_A as the "column space" and R_{A^*} as the "row space" (if A has <u>real</u> entries).

<u>Theorem 8.6</u>: If $N: X \to X$ is normal, then X has an orthonormal basis of eigenvectors of N.

<u>Proof</u>: Write $N = H + A$, where $H = \frac{N+N^*}{2}$, $A = \frac{N-N^*}{2}$.

If N & N^* commute, then H and iA commute, and these are self-adjoint anyways.

By Theorem 8.4, they have a common spectral resolution, thus X has an orthonormal basis of common eigenvectors.

However, since $N = H + A$, these are eigenvectors of N (and N^*) as well. $\square$

<u>Theorem 8.7</u>: Let $U: X \to X$ be unitary. Then

(a) X has an orthonormal basis of eigenvectors of U.

(b) Each eigenvalue has norm 1.

<u>Proof</u>: (a) Immediate from Theorem 8.6.

(b) If $Uv = \lambda v$, then $\|Uv\| = \|v\|$ since U is unitary

$$\Rightarrow \|Uv\| = \|\lambda v\| = |\lambda| \cdot \|v\| = \|v\| \quad \Rightarrow \quad |\lambda| = 1. \quad \square$$

Recall that we derived the spectral resolution of self-adjoint maps using the spectral theory of general maps. Here, we'll give an alternate proof that has several advantages:

- It doesn't assume the fundamental theorem of algebra.
- For real symmetric matrices, it avoids complex numbers.
- It leads to the "minmax principle" which gives a new characterization of the eigenvalues of H. (And other applications!)

First, suppose X has an orthonormal basis of eigenvectors of a mapping $M: X \to X$ and write $x = (a_1, \dots, a_n)$ in this basis.

Define: • $q(x) := (x, Mx) = \left(\sum_{i=1}^{n} a_i v_i , \sum_{i=1}^{n} a_i M v_i \right)$

$$= \left(\sum_{i=1}^{n} a_i v_i , \sum_{i=1}^{n} a_i \lambda_i v_i \right) = \sum_{i=1}^{n} \lambda_i a_i^2 .$$

• $p(x) = (x, x) = \sum_{i=1}^{n} a_i^2 .$

Def: Let $H: X \to X$ be self-adjoint and define the Rayleigh quotient of H by $R(x) = R_H(x) = \dfrac{(x, Hx)}{(x, x)}$.

Goal: Show that the minimum & maximum values of $R(H)$ (and actually, all critical points!) occur at the eigenvectors of H. Deduce that H has a full set of orthonormal eigenvectors.

Remark: Since $R(kx) = R(x)$, we only need to consider unit vectors.

Suppose that $R(v) = \min\{R(x) : \|x\| = 1\} := \lambda$. [and $\|v\| = 1$]

Let $w \in X$ be any other vector, and $t \in \mathbb{R}$ a parameter.

$$R(v+tw) = \frac{(v+tw, H(v+tw))}{(v+tw, v+tw)} = ($$

$$= \frac{(v, Hv) + t(v, Hw) + t(w, Hv) + t^2(w, w)}{(v, v) + t(v, w) + t(w, v) + t^2(w, w)}$$

$$= \frac{(v, Hv) + 2t\,\mathrm{Re}(Hv, w) + t^2(w, w)}{(v, v) + 2t\,\mathrm{Re}(v, w) + t^2(w, w)} = \frac{q(t)}{p(t)}.$$

Since R is minimized at $t=0$, we know that

$$\dot{R}(0) = \frac{d}{dt}\left(\frac{q(t)}{p(t)}\right)\bigg|_{t=0} = \frac{p(0)\,\dot{q}(0) - \dot{p}(0)\,q(0)}{(p(0))^2} = 0$$

At $t=0$: $\quad p(0) = (v,v) = 1 \qquad q(0) = R(v) = \lambda$

$\dot{p}(0) = 2\,\mathrm{Re}(v,w) \qquad \dot{q}(0) = 2\,\mathrm{Re}(Hv, w).$

$$\Rightarrow p(0)\dot{q}(0) - \dot{p}(0)q(0) = 1\cdot 2\,\mathrm{Re}(Hv,w) - \lambda\cdot 2\,\mathrm{Re}(v,w)$$

$$= 2\,\mathrm{Re}(Hv-\lambda v, w) = 0 \qquad \forall w\in X$$

Since this holds for all $w\in X$, $Hv - \lambda v = 0 \Rightarrow Hv = \lambda v$.

Now, let $X_1 = \mathrm{Span}(v)^{\perp}$, so $X = X_1 \oplus \mathrm{Span}(v)$ and $\dim X_1 = n-1$.

<u>Claim</u>: X_1 is "H-invariant"; that is, H maps X_1 into X_1

<u>Proof</u>: $(x,v) = 0 \Rightarrow (Hx, v) = (x, Hv) = (x, \lambda v) = \lambda(x,v) = 0.$

That is, if $x\in X_1$, then $Hx \in X_1$.

Now, put $v_1 = v$ and $\lambda_1 = \lambda$.

Let $v_2 \in X_1$ be the (nonzero) vector for which

$$R(v_2) = \min\{R(x) : x\in X_1,\ \|x\| = 1\} := \lambda_2$$

Then v_2 is an eigenvector of H with eigenvalue $\lambda_2 \geq \lambda_1$

Next, put $X_2 := \mathrm{Span}(x_1, x_2)^{\perp}$ and continue in this fashion

We get a full set of orthonormal eigenvectors of H with eigenvalues $\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$.

<u>Theorem 8.8</u>: (Minmax principle). Let $H: X \to X$ be self-adjoint with eigenvalues $\lambda_1 \le \dots \le \lambda_n$. Then $\lambda_k = \min_{\dim S = k} \left\{ \max_{x \in S \setminus \vec{0}} R_H(x) \right\}$.

<u>Proof</u>: Let S be any k-dimensional subspace.

First, we'll show that $R_H(x) \geq \lambda_k$ for some $x \in S$.

Let $v_1, \dots, v_n$ be the eigenvectors, assume $\|v_i\| = 1$

Let $T = \operatorname{Span}\{v_k, \dots, v_n\}$ so $\dim T = n-(k-1) = n-k+1$

Thus, $\dim S + \dim T - \dim S \cap T = \dim S+T \le n$

$\Rightarrow \quad k + (n-k+1) - d \le n$

$\Rightarrow \quad d \geq 1$,

Thus there is some $x \in S \cap T$, $\|x\| = 1$.

Write $x = \sum_k^n a_i v_i \Rightarrow R(x) = (x, Hx) =$

$\Rightarrow R(x) = (x, Hx) = \sum_{i=k}^n \lambda_i a_i^2 \geq \lambda_k \sum_{i=k}^n a_i^2 = \lambda_k$ ✓

Next, show that some k-dimensional subspace achieves this minimum, i.e., find $S \subseteq X$ for which $R(x) \leq \lambda_k$ for all $x \in S$.

Take $S = \mathrm{Span}\{v_1, \dots, v_k\}$.

For any unit vector $x = \sum_{i=1}^{k} b_i v_i \in S$,

$$R(x) = (x, Hx) = \sum_{i=1}^{k} \lambda_i b_i^2 \leq \lambda_k \sum_{i=1}^{k} b_i^2 = \lambda_k. \quad ✓ \qquad \square$$

Summary of the Rayleigh quotient:

(i) Every eigenvector v_i of H is a critical point of $R_H(x)$, i.e., the 1st derivatives of $R_H(x)$ are zero iff x is an eigenvector

(ii) For any eigenvector v_i with eigenvalue λ_i, $R_H(v_i) = \lambda_i$.

(iii) In particular, $\lambda_1 = \min\{R(x) : x \neq 0\}$

$$\lambda_n = \max\{R(x) : x \neq 0\}.$$

Application: Let H be real-symmetric, and let v be an eigenvector with eigenvalue λ. If $\|v - w\| \leq \varepsilon$, then $\|v - R_H(w)\| \leq O(\varepsilon^2)$, i.e., $R_H(w)$ is a 2nd order Taylor approximation of the eigenvalue. This arises in numerical methods for computing eigenvalues.

<u>Def</u>: A self-adjoint map $M: X \to X$ is <u>positive</u> (or <u>positive definite</u>) if $(x, Mx) > 0$ for all $x \neq 0$.

<u>Remark</u>: From our analysis of the Rayleigh quotient, M is positive iff all eigenvalues of M are positive.

<u>Generalized Raleigh quotient</u>: If $H, M: X \to X$ are self-adjoint and M positive, the define $R_{H,M}(x) = \dfrac{(x, Hx)}{(x, Mx)}$.

Note that $R_H = R_{H,I}$.

We can derive a similar minmax principle:

<u>Theorem 8.9</u>: The minimum problem $\min\{R_{H,M}(x)\}$ has a solution $R_{H,M}(v) = \mu$, where $v \neq 0$ and μ solves $Hv = \mu Mv$. The (constrained) minimum problem $\min\{R_{H,M}(x) : (x, Mv) = 0\}$ has a solution $R_{H,M}(w) = \nu$ where $w \neq 0$ and ν satisfies $Hw = \nu Mw$.

<u>Proof</u>: Exercise. (HW)

As before, we can iterate this process and produce a special basis for X

Theorem 8.10: Let $H, M : X \to X$ be self-adjoint and M positive. Then there is a basis $v_1, \dots, v_n$ of X where each v_i satisfies $Hv_i = \mu_i M v_i$ for some $\mu_i \in \mathbb{R}$, and $(v_i, Mv_j) = 0$ for $i \neq j$.

Corollary 8.11: All eigenvalues of $M^{-1}H$ are real. Moreover, if H is also positive, then the eigenvalues of $M^{-1}H$ are all positive

Proof: Exercise (HW).

Theorem 8.12: Let $N : X \to X$ be a normal linear map. Then $\|N\| = \max |\lambda_i|$, taken over all eigenvalues of N.

Proof: Exercise (HW).

Recall that for any linear map $A : X \to U$, the matrix $A^*A : X \to X$ is self-adjoint and nonnegative (that is, $(x, Mx) \geq 0$ $\forall x \in X$.) It is positive if $N_A = \{0\}$, (because rank A = rank A^*A.)

Thus, in some sense, the matrix A^*A is the "proper" way to think of the "square" of a matrix.

[Note: In contrast, A^2 could have negative eigenvalues.]

The next result even further supports this claim:

Theorem 8.13: Let $A: X \to X$ be linear and say that the eigenvalues of A^*A are $\lambda_1 \le \cdots \le \lambda_n$. Then $\|A\| = \sqrt{\lambda_n}$.

Proof: We need to show $\max\{\|Ax\|^2 : \|x\| = 1\} = \lambda_k$.

First take any $x \in X$ with $\|x\| = 1$:

$$\|Ax\|^2 = (Ax, Ax) = (x, A^*Ax) \underset{\text{Cauchy-Schwarz}}{\le} \|x\| \cdot \|A^*Ax\| = \|A^*Ax\| \underset{\text{Theorem 8.12}}{\le} \lambda_n$$

Thus, $\|Ax\| \le \sqrt{\lambda_n}$.

To show equality, it suffices to find some $x \in X$, $\|x\| = 1$ for which $\|Ax\| = \sqrt{\lambda_n}$.

Take the corresponding eigenvector v_n of A^*A:

$$\|Av_n\|^2 = (Av_n, Av_n) = (v_n, A^*Av_n) = (v_n, \lambda_n v_n) = v_n . \checkmark \qquad \square$$