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

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 2nd 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 $g \in \mathbb{R}$?

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

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

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

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

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

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

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

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 $(v,w) = 0$. ✓

□

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

By Theorem 8.1, we can write $X = N^{(1)} \oplus \cdots \oplus N^{(k)}$, 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 $x = x^{(1)} + \cdots + x^{(k)}$, $x^{(i)} \in N^{(i)}$.

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

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

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

<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, $H : X \to X$ a self-adjoint linear map. Then there is a resolution of the identity and a spectral resolution of H (in the sense described above).

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

$H^2 = \sum_{i=1}^{k} \lambda_i^2 P_i$, $H^m = \sum_{i=1}^{k} \lambda_i^m P_i$, 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>: $e^H = \sum_{i=1}^{k} e^{\lambda_i} P_i$.

<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 $X = N^{(1)} \oplus \dots \oplus N^{(k)}$, a product of eigenspaces of H corresponding to distinct eigenvectors. Pick $N = N^{(i)}$. 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 $N \longrightarrow N$. Apply a spectral resolution to K over N, i.e, write

$$K|_N = \sum_{i=1}^{k_i} \mu_i P_i \quad \text{and} \quad H|_N = \sum_{i=1}^{h_i} \lambda_i P_i, \quad I|_N = \sum P_i.$$

The union of such a decomposition for all $N^{(i)}$ is a common spectral resolution of H and K, which we seek. $\square$

<u>Remark</u>: This is easily generalized for any number of commuting maps.

Note that $(iM)^* = -iM^*$ (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, from Theorem 8.1:

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.

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

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

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

Proof: Write $N = H + A$, where $H = \frac{N+N^*}{2}$, $A = \frac{N-N^*}{2}$.
If N and N^* commute, then H and iA commute and are self-adjoint. By Theorem 8.6, 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. $\square$

Theorem 8.7: Let $U: X \to X$ be a unitary map, Then

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

(b) Each eigenvalue has norm 1.

<u>Proof:</u> (a) is immediate from the fact that U is normal.

For (b), if $Uv = \lambda v$, then

$$\|Uv\| = \|\lambda v\| = |\lambda| \, \|v\| \Rightarrow |\lambda| = 1. \quad \checkmark \qquad \square$$

Recall that we proved 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.

First, suppose X has an orthonormal basis of eigenvectors of a mapping $M : X \to X$, and write $x = (z_1, \dots, z_n)$ in this basis.

We can write $q(x) := (x, Mx) = \sum_{i=1}^{n} \lambda_i z_i$.

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

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

Goal: Show that the minimum (and maximum) values of $R(H)$ are taken at eigenvectors of H. Deduce that H has a full set of orthonormal eigenvectors.

Approach: Since $R(kx) = k R(x)$, we need only to consider unit vectors. Suppose that $R(v) = \min\{R(x) : \|x\| = 1\} = \lambda$. Let $w \in X$ be any other vector, and $t \in \mathbb{R}$ a parameter.

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

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

$$\dot{R} := \frac{d}{dt} R(v+tw)\Big|_{t=0} = \frac{d}{dt}\left(\frac{q(t)}{p(t)}\right)\Big|_{t=0} = \frac{p(0)\,\dot{q}(0) - \dot{p}(0)\,q(0)}{(p(0))^2} = 0.$$

At $t=0$, $\quad p(0) = (v,v) = 1, \quad \dot{p}(0) = 2\,\mathrm{Re}(w,v)$

$$q(0) = R(v) = \lambda, \quad \dot{q}(0) = 2\,\mathrm{Re}(w, Hv)$$

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

$$\Rightarrow 2\,\mathrm{Re}(w, Hv - \lambda v) = 0.$$

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

Now, write $X_1 = \langle v \rangle^{\perp}$, so $X = X_1 \oplus \langle v \rangle$, $\dim X_1 = n-1$.

Remark: H maps $X_1 \to X_1$ because for any $x \in X_1$,

$(x, v) = 0 \Rightarrow (Hx, v) = (x, Hv) = (x, \lambda v) = \lambda(x, v) = 0.$

Put $v_1 = v$ and $\lambda_1 = \lambda$. Let $v_2 \in X_1$ be the vector for which

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

v_2 is an eigenvector of H which eigenvalue $\lambda_2 \geq \lambda_1$.

Next, put $X_2 = \langle v_1, v_2 \rangle^\perp$, and continue in this fashion.

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

<u>Theorem 8.8</u> (Minmax principle): Let $H : X \to X$ be a self-adjoint map of a finite-dimensional real Euclidean space, with eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$. Then $\lambda_k = \min_{\dim S = k} \left\{ \max_{x \in S \setminus 0} R_H(x) \right\}$.

<u>Proof</u>: Let $S \subseteq X$ be a k-dimensional subspace.

<u>Goal</u>: First show that $\max\{R_H(x) : x \in S \setminus 0\} \geq \lambda_k$.

Consider the (underdetermined) system of equations

$(x, v_i) = 0 \qquad i = 1, \dots, k-1$

Let v be a solution (assume $\|v\| = 1$).

Write $v = a_1 v_1 + \dots + a_n v_n$. ($v_1, \dots, v_n$ are orthonormal eigenvectors)

Clearly, $a_1 = \dots = a_{k-1} = 0$.

$$R(v) = \frac{(v, Hv)}{(v,v)} = \frac{\sum_{i=1}^{n} \lambda_i a_i^2}{\sum_{i=1}^{n} a_i^2} = \sum_{i=k}^{n} \lambda_i a_i \geq \lambda_k \sum_{i=1}^{n} a_i^2 = \lambda_k \quad ✓$$

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 $S = \langle v_1, \ldots, v_k \rangle$.

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

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

$\square$

In summary, we have shown the following properties of the Rayleigh quotient:

(i) Every eigenvector v_i of H is a critical point of $R_H(x)$, i.e, the 1^{st} 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, $\lambda_1 = \min\{R(x) : x \neq 0\}$.

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

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

<u>Def:</u> A self-adjoint mapping $M : X \to X$ is <u>positive definite</u> if $(x, Mx) > 0$ for all $x \neq 0$ (we'll study this more later.)

<u>Generalized Rayleigh quotient</u>: If $H, M : X \to X$ are self-adjoint and M positive definite, then define

$$R_{H,M}(x) = \frac{(x, Hx)}{(x, Mx)}. \qquad (\text{Note that } R_H = R_{H,I}.)$$

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 solve $Hv = \mu Mv$. 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 satisfy $Hw = \nu Mw$.

<u>Proof:</u> Exercise.

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

<u>Theorem 8.10</u>: Let $H, M : X \to X$ be self-adjoint and M positive definite. Then there exists 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 $i \neq j$.

<u>Corollary 8.11</u>: All the eigenvalues of $M^{-1}H$ are real. Moreover, if M is positive definite, then the eigenvalues of $M^{-1}H$ are positive.

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

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

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

<u>Theorem 8.13</u>: Let $A : X \to X$ be a linear map of a finite-dimensional Euclidean space, with eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$. Then $\|A\| = \sqrt{\lambda_n}$.

<u>Proof</u>: $\|Ax\|^2 = (Ax, Ax) = (x, A^*Ax) \leq \|x\| \, \|A^*Ax\|$, by Cauchy-Schwarz. Thus for any unit vector $x \in X$, $\|Ax\|^2 \leq \|A^*Ax\|$.