8. Self-adjoint mappings:

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

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

\[
M = H + A, \quad H = \frac{M + M^*}{2}, \quad A = \frac{M - M^*}{2}.
\]

Indeed,

\[
\text{Re}(x, Mx) = \frac{1}{2} \left[ (x, Mx) + (x, M^*x) \right] = \frac{1}{2} \left[ (x, M^*x) + (x, Mx) \right] = \frac{1}{2} \left[ (x, Mx) + (x, M^*x) \right] = (x, Hx)
\]

\[
\text{Im}(x, Mx) = \frac{1}{2} \left[ (x, Mx) - (x, M^*x) \right] = \frac{1}{2} \left[ (x, M^*x) - (x, Mx) \right] = \frac{1}{2} \left[ (x, M^*x) - (x, M^*x) \right] = (x, Ax).
\]

**Quadratic forms**

**Motivation:** Let \( f(x_1, \ldots, x_n) \) be a real-valued function, \( \mathbb{R}^n \to \mathbb{R} \).

Recall the **Taylor approximation** of \( f \) at \( a \in \mathbb{R}^n \) up to 2nd order says that, for \( y \in \mathbb{R}^n \) with \( \|y\| \approx 0 \),

\[
f(a+y) \approx f(a) + l(y) + \frac{1}{2} q(y),
\]

where
* $f(a)$ is the 0th order term

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

It turns out that $g = \nabla f = \left( \frac{df}{dx_1}, \ldots, \frac{df}{dx_n} \right)$, the gradient of $f$.

* $g(y)$ is the 2nd order term: $g(y) = \sum_{j=1}^{n} \sum_{i=1}^{n} h_{ij} y_i y_j$, where $H = (h_{ij}) = \left( \frac{d^2f}{dx_i dx_j} \right)$ is the Hessian of $f$.

Note that $H$ is self-adjoint, because $\frac{d^2f}{dx_i dx_j} = \frac{d^2f}{dx_j dx_i}$.

and that $g(y) = [y_1, \ldots, y_n] H [y_1, \ldots, y_n]^T = (y, Hy)$.

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 2nd order term $g(y)$.

**Def:** A function $g : X \to \mathbb{K}$ of the form $g(x) = (x, Hx)$ for a self-adjoint map $H$ is called a quadratic form.

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

Suppose now that we can diagonalize $H$, that is, write $H = P^{-1}DP$. Recall that this would mean that $D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$ 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
\[ q(z) = z^T D z = [z_1, \ldots, z_n] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \sum_{i=1}^{n} \lambda_i z_i^2. \]
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.

**Theorem 8.1:** 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 \( X \).

**Proof:** 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.

**Proof:** (i) Let \( \lambda \) be an eigenvalue of \( H \) with eigenvector \( v \neq 0 \).
Then \( (Hv, v) = (\lambda v, v) = \lambda (v, v) \)
and \( (v, Hv) = (v, \lambda v) = \overline{\lambda} (v, v) \)
Since \((v, v) \neq 0\), \(\lambda = \bar{\lambda} \Rightarrow \lambda \) is real. \(\checkmark\)

(ii) Suppose \((H - \lambda I)^d v = 0\). We must show \((H - \lambda I)^{d-1} v = 0\).

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

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

\[ \Rightarrow \quad \langle (H - \lambda I)v, (H - \lambda I)v \rangle = \| (H - \lambda I)v \|^2 = 0 \Rightarrow (H - \lambda I)v = 0. \checkmark \]

Now, suppose \((H - \lambda I)^d v = 0 \Rightarrow (H - \lambda I)^2 (H - \lambda I)^{d-2} v = 0\)

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

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

\[ \Rightarrow \quad (H - \lambda I)v = 0 \quad \text{(induction hypothesis)} \checkmark \]

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

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

So if \(\lambda \neq \mu\), then \(\langle v, w \rangle = 0\). \(\checkmark\)

\[ \square \]

Corollary 8.2: If \(H\) is self-adjoint, then \(H = M^* D M\) 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 \(\lambda_i\), 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 projection of $x$ onto the eigenspace $N_i$, that is

$$P_i : X \to X, \quad P_i : x \mapsto x_i.$$

**Remark:**
(a) $P_i P_j = 0$ if $i \neq j$ and $P_i^2 = P_i$

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

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

**Corollary 8.2** can now be stated as follows:

**Theorem 8.3:** 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 $H$.

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

$$H^2 = \sum_{i=1}^k \lambda_i^2 P_i, \quad H^n = \sum_{i=1}^k \lambda_i^n P_i,$$

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

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

**Theorem 8.1:** 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 \( I = \sum_{i=1} P_i \) and \( H = \sum_{i=1} \lambda_i P_i \) and \( K = \sum_{i=1} \mu_i P_i \).

**Proof:** Write \( X = N(1) \oplus \ldots \oplus N(k) \), a product of eigenspaces of \( H \) corresponding to distinct eigenvectors.

Pick \( N = N(1) \). Then for every \( x \in N \), \( Hx = \lambda x \)

\[ H(Kx) = K(Hx) = K(\lambda x) = \lambda (Kx) \]

Thus, \( Kx \) is an eigenvector of \( H \), so \( K \) maps \( N \to N \).

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

\[ K|_N = \sum_{i=1}^{k_j} \mu_{ji} P_{ji} \quad \text{and} \quad I|_N = \sum_{i=1}^{k_j} P_{ji}. \]

Assume \( \mu_i \)'s distinct.

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

Now, \( N(i) = N(i1) \oplus N(i2) \oplus \ldots \oplus N(i(k_i)) \), orthogonal eigenspace of \( K|_N \) (and of \( H|_N \)?)
Expanding each $N^{(i)}$ into eigenspace of $K/W$ gives a common spectral resolution of $H$ and $K$, which we seek.

That is, 

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

$$= (N^{(1)}) \oplus \ldots \oplus N^{(k_1)} \oplus (N^{(2)}) \oplus \ldots \oplus N^{(k_2)} \oplus \ldots \oplus (N^{(k)}) \oplus \ldots \oplus N^{(k_{k_2})}$$

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

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 $A$. 

Def: A mapping \( N : X \to \mathbb{X} \) 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.

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

\[
\begin{align*}
&X \\
&\quad R_{A^*} \\
&\quad \dim n-r \\
&\quad N_A \\
&\quad \dim n-r \\
\end{align*}
\] 

\[
\begin{align*}
&U \\
&\quad R_A \\
&\quad \dim r \\
&\quad N_{A^*} \\
&\quad \dim m-r \\
\end{align*}
\]

Facts (proofs are HW):

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

\[
\text{and so } X = R_{A^*} \oplus N_A \text{ and } U = R_A \oplus N_{A^*}.
\]

Think of \( R_A \) as the "column space" and \( R_{A^*} \) as the "row space" (if \( A \) has real entries).
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 there 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. \( \square \)

Theorem 8.7: Let \( U : X \to X \) be unitary. Then

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

(b) Each eigenvalue has norm 1.

Proof: (a) Immediate from Theorem 8.6.

(b) If \( UV = \lambda U \), then \( \|UV\| = \|V\| \) since \( U \) is unitary.

\[ \Rightarrow \|UV\| = \|\lambda U\| = |\lambda| \|U\| = \|V\| \Rightarrow |\lambda| = 1. \square \]
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 $x = (a_1, \ldots, a_n)$ in this basis.

Define:
- $q(x) = (x, Mx) = \left( \sum_{i=1}^{\hat{n}} a_i v_i, \sum_{i=1}^{\hat{n}} a_i M v_i \right)$
  
  $\quad = \left( \sum_{i=1}^{\hat{n}} a_i v_i, \sum_{i=1}^{\hat{n}} a_i \lambda_i v_i \right) = \sum_{i=1}^{\hat{n}} \lambda_i a_i^2$.

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

Define: let $H: X \to X$ be self-adjoint and define the Rayleigh quotient of $H$ by $R(x) = R_H(x) = \frac{(x, Hx)}{(x, x)}$. 
Goal: Show that the minimum & maximum values of $R(t)$
(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 $w \neq x$ be any other vector, and $t \in \mathbb{R}$ a parameter.

\[
R(v+tw) = \frac{(v+tw, H(v+tw))}{(v+tw, v+tw)} = \left( \frac{(v, Hv) + t(v, Hw) + t^2(w, w)}{(v, v) + t(v, w) + t^2(w, w)} \right)
\]

\[
= \frac{(v, Hv) + 2t \text{Re}(Hv, w) + t^2(w, w)}{(v, v) + 2t \text{Re}(v, w) + t^2(w, w)} = \frac{\Phi(t)}{\rho(t)}.
\]

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

\[
\dot{R}(0) = \left. \frac{d}{dt} \frac{\Phi(t)}{\rho(t)} \right|_{t=0} = \frac{\rho(0) \dot{\Phi}(0) - \Phi(0) \dot{\rho}(0)}{(\rho(0))^2} = 0
\]
At \( t=0 \):

\[ p(0) = (\nu, \nu) = 1 \quad \quad q(0) = R(\nu) = \lambda \]

\[ \dot{p}(0) = 2 \text{ Re} (\nu, w) \quad \quad \dot{q}(0) = 2 \text{ Re} (H\nu, w). \]

\[
\Rightarrow \quad p(0) \dot{q}(0) - \dot{p}(0) q(0) = 1 \cdot 2 \text{ Re} (H\nu, w) - \lambda \cdot 2 \text{ Re} (\nu, w) \\
= 2 \text{ Re} (H\nu - \lambda \nu, w) = 0 \quad \forall \nu \in X
\]

Since this holds for all \( \nu \in X \), \( H\nu - \lambda \nu = 0 \Rightarrow H\nu = \lambda \nu. \)

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

Claim: \( X_1 \) is "\( H \)-invariant"; that is, \( H \) maps \( X_1 \) onto \( X_1 \).

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

That is, if \( x \in X_1 \), then \( Hx \in X_1 \).

Now, put \( v_1 = \nu \) and \( \lambda_1 = \lambda \).

Let \( v_2 \in X \), be the (non-zero) vector for which

\[ R(v_2) = \min \left\{ R(x) : x \in X_1, \|x\| = 1 \right\} =: \lambda_2 \]

Then \( v_2 \) is an eigenvector of \( H \) with eigenvalue \( \lambda_2 \geq \lambda \).

Next, put \( X_2 := \text{Span} (v_1, v_2)^\perp \) and continue in this fashion.
We get a full set of orthonormal eigenvectors of $H$ with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

**Theorem 8.8:** (Min-max principle). Let $H: X \to X$ be self-adjoint with eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$. Then $\lambda_k = \min_{\dim S = k} \{ \max_{x \in S, \|x\| = 1} R_H(x) \}$.

**Proof:** Let $S$ be any $k$-dimensional subspace.

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

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

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

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

$\implies k + (n - k + 1) - d \leq n$

$\implies d \geq 1$.

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

Write $x = \sum_{i=1}^{n} a_i v_i \implies R(x) = (x, Hx) =\sum_{i=1}^{n} a_i^2 \lambda_i$.

$\implies R(x) = \langle x, Hx \rangle = \sum_{i=k}^{n} \lambda_i a_i^2 \geq \lambda_k \sum_{i=k}^{n} a_i^2 = \lambda_k$.
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 $S = \text{Span}\{v_1, \ldots, v_k\}$.

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

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

Summary of the Rayleigh quotient:

(i) Every eigenvector $v_i$ of $H$ is a critical point of $\hat{R}_H(x)$, i.e., the $1^{st}$ derivatives of $\hat{R}_H(x)$ are zero iif $x$ is an eigenvector.

(ii) For any eigenvector $v_i$ with eigenvalue $\lambda_i$, $\hat{R}_H(v_i) = \lambda_i$.

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

$$\lambda_1 = \max \{R(x) : x \neq 0\}.$$  

Application: Let $H$ be real-symmetric, and let $v$ be an eigenvector with eigenvalue $\lambda$. If $\|v-w\| \leq \varepsilon$, then $\|v - \hat{R}_H(w)\| = O(\varepsilon^2)$, i.e., $\hat{R}_H(w)$ is a 2nd order Taylor approximation of the eigenvalue. This arises in numerical methods for computing eigenvalues.
**Def.** A self-adjoint map $M: X \to X$ is **positive** (or **positive definite** if $(x, Mx) > 0$ for all $x \neq 0$.

**Remark.** From our analysis of the Rayleigh quotient, $M$ is positive iff all eigenvalues of $M$ are positive.

**Generalized Rayleigh quotient:** If $H, M: X \to X$ are self-adjoint and $M$ positive, then define $R_{H, M}(x) = \frac{(x, Hx)}{(x, Mx)}$.

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

We can derive a similar minmax principle:

**Theorem 8.9:** The minimum problem $\min \{ R_{H, M}(x) \}$ has a solution $R_{H, M}(w) = \mu > 0$ where $w \neq 0$ and $\mu$ solves $Hv = \mu Mv$.

The (constrained) minimum problem $\min \{ R_{H, M}(x) : (x, Mw) = 0 \}$ has a solution $R_{H, M}(w) = \nu$ where $w \neq 0$ and $\nu$ satisfies $Hw = \nu Mw$.

**Proof.** 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 \( v_1, \ldots, v_n \) of \( X \) where each \( v_i \) satisfies \( Hv_i = \mu_i Mv_i \) 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 non-negative (that is, \( (x, Mv) \geq 0 \) \( \forall x \in X \)). It is positive if \( N_A = \{0\} \), (because \( \text{rank } A = \text{rank } A^*A \)).
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 $A : X \rightarrow X$ be linear and say that the eigenvalues of $A^*A$ are $\lambda_1 \leq \cdots \leq \lambda_n$. Then $\|A\| = \sqrt{\lambda_n}$.

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

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

$$\|Ax\|^2 = (Ax, Ax) = (x, A^*Ax) \leq \|x\| \cdot \|A^*Ax\| = \|A^*Ax\| \leq \lambda_n$$

Cauchy-Schwarz

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

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 $A^*A$:

$$\|AV_n\|^2 = (AV_n, AV_n) = (V_n, A^*AV_n) = (V_n, \lambda_n V_n) = \lambda_n V_n.$$

Thus, $\|AV_n\| = \sqrt{\lambda_n}$. 

$\Box$