

Lecture 21: Diagonalizable matrices

Reading: Meyer 7.5 Normal matrices
7.6 Positive semi-definite matrices

SPECTRAL THEORY

Questions:

- ① What matrices have an eigenvector?
- ② What matrices can be diagonalized?
(ie, they have a full basis of eigenvectors, $A = UDU'$)
- ③ Which can be diagonalized with a basis of orthogonal eigenvectors?

Answers

all square matrices

diagonalizable matrices
including matrices with n distinct eigenvalues
(multiplicity 1)

? TODAY

THE "POWER METHOD" FOR FINDING THE LARGEST EIGENVALUE EIGENVECTOR

Matlab `eig()` and `eigs()` commands

```
>> help eig
eig Eigenvalues and eigenvectors.
E = eig(A) produces a column vector E containing the eigenvalues of
a square matrix A.

[V,D] = eig(A) produces a diagonal matrix D of eigenvalues and
a full matrix V whose columns are the corresponding eigenvectors
so that A*V = V*D.

>> help eigs
eigs Find a few eigenvalues and eigenvectors of a matrix
D = eigs(A) returns a vector of A's 6 largest magnitude eigenvalues.
A must be square and should be large and sparse.
```

```
>> help sprandsym
sprandsym Sparse random symmetric matrix.
R = sprandsym(S) is a symmetric random matrix whose lower triangle
and diagonal have the same structure as S. The elements are
normally distributed, with mean 0 and variance 1.

R = sprandsym(n,density) is a symmetric random, n-by-n, sparse
matrix with approximately density*n*n nonzeros; each entry is
the sum of one or more normally distributed random samples.
```

Example:

```
>> n = 2000; (roughly 10 nonzero entries per row)
>> density = 10/n;
>> A = sprandsym(n, density);
>> tic; eig(A); toc
Elapsed time is 18.907694 seconds.
```

find all e-values

find largest-magnitude e-value

Elapsed time is 18.907694 seconds.

find all e-values

Elapsed time is 32.109028 seconds.

find largest-magnitude e-value

How this works

Let A be a diagonalizable matrix (for simplicity).

Sort its eigenvalues by magnitude:

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|.$$

Algorithm (simplest version)

- ① Start with a "generic" vector \vec{x}_0
- ② Repeat for $t = 1, 2, 3, \dots$
let $\vec{x}_t = A\vec{x}_{t-1}$

Analysis:

Expand \vec{x}_0 in the basis of eigenvectors $\vec{v}_1, \dots, \vec{v}_n$:

$$\vec{x}_0 = \sum_{j=1}^n c_j \cdot \vec{v}_j$$

$$\Rightarrow A\vec{x}_0 = \sum_j c_j \lambda_j \vec{v}_j$$

$$\Rightarrow A^t \vec{x}_0 = \sum_j c_j \lambda_j^t \vec{v}_j$$

↑
this is dominated by the
largest λ_j 's

$$= \lambda_1^t \cdot \left(c_1 \vec{v}_1 + \sum_{j \geq 1} c_j \left(\frac{\lambda_j}{\lambda_1} \right)^t \vec{v}_j \right)$$

Provided $c_1 > 0$ and $|\lambda_2| < |\lambda_1|$, this will converge exponentially quickly to a multiple of \vec{v}_1 .

It converges faster if $|\lambda_2/\lambda_1|$ is smaller.

Problem: What if $c_1 = 0$?

This shouldn't happen.

Problem: Numerical overflow

Problem: Numerical overflow

```
>> n = 3;
>> A = randn(n, n);
>>
>> v = randn(n, 1);
>> for t = 1:10000
    v = A * v;
end
```

>> v

v =

NaN
NaN
NaN

numbers got
too big!

Solution:
renormalize
occasionally

```
>> v = randn(n, 1);
>> for t = 1:10000
    v = A * v;
    v = v / norm(v);
end
```

>> v

v =
0.3950
0.7841
0.4786

>> (A * v) ./ v

ans =

1.1099
1.1099
1.1099

Problem: What if $|\lambda_1| = |\lambda_2|$?

Example: Permutation matrix

>> A = [0 1 0; 0 0 1; 1 0 0]

A =

0	1	0
0	0	1
1	0	0

```
>> v = randn(3, 1);
>> for t = 1:10000
    v = A * v;
end
```

>> v

v = ← not an e-vector!

0.3426
3.5784
-0.4336

>> Ashift = A + .01*randn() * eye(3)

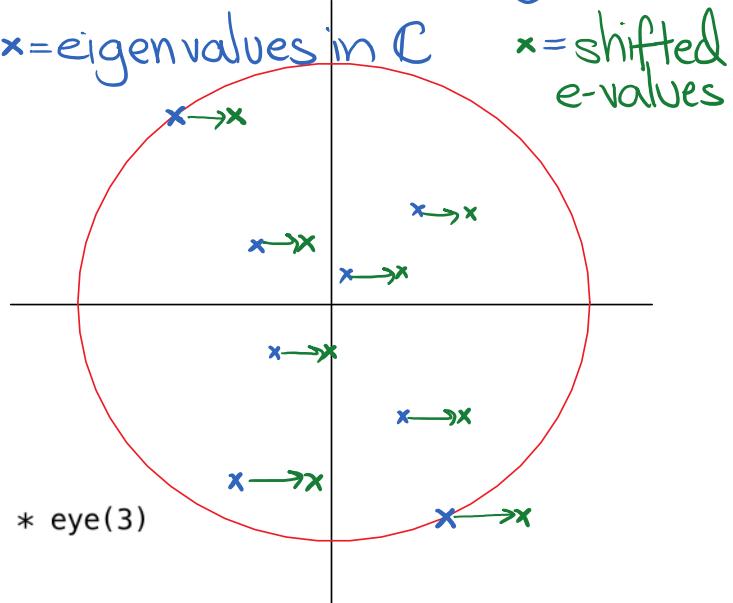
Ashift =

0.0163	1.0000	0
0	0.0163	1.0000
1.0000	0	0.0163

```
>> v = randn(3, 1);
>> for t = 1:10000
    v = Ashift * v;
    v = v / v(1); different way to
    end prevent blowup
```

Solution: Break the degeneracy

x=eigenvalues in C x=shifted e-values



```

v = v / v(1); different way to
end prevent blowup
>> v

```

v =

1.0000
1.0000
1.0000

This works more reliably
if the e-values are all real

Problem: What about other e-values?

We'll see...

Smallest magnitude e-value

eigs(A, 1, 'sm') finds smallest magnitude

Exercise: How does this work?

for t=1,2,...

$\vec{v} = A \setminus \vec{v}$; ← don't do $\vec{v} \leftarrow A^{-1} \vec{v}$

$\vec{v} = \vec{v} / \text{norm}(v)$; ← precomputing the LU
decomposition will be faster

Exercise: Using Matlab, find the e-value closest to 10.

Answer 1: eigs(A-10*eye(n), 1, 'sm')

Answer 2: Using the power method:

B = A - 10 * eye(n);

x = randn(n, 1);

for j=1:10000 ~ it would be better to check
for convergence...

x = B \ x;

x = x / norm(x); ← compute $B^{-1}x$

end

it would be faster to precompute the
LU decomposition of B

WHEN IS A MATRIX DIAGONALIZABLE?

When does it have a complete set of eigenvectors?

Exercise: Which of these matrices can be diagonalized?

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Answer:

A and C.

A: Since A is triangular, you can read its eigenvalues off the diagonal: A's eigenvalues are 1 and 2.

Two different e-values \Rightarrow two independent e-vectors, and in \mathbb{R}^2 that's all there's room for. ✓

B: B has eigenvalue 1 with multiplicity 2, but the associated eigenspace $N(B - I) = N\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is only one-dimensional. B does not have a complete set of e-vectors.

C: $C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. It is proportional to the projection $\vec{v}\vec{v}^T$ for $\vec{v} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The orthogonal direction is $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Eigenvector	Eigenvalue
$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	2
$\begin{pmatrix} -1 \\ 1 \end{pmatrix}$	0

A matrix is diagonalizable when

- for each eigenvalue, the dimension of the associated eigenspace equals the multiplicity of the eigenvalue.
- In other words, if

$$\det(A - \lambda I) = (\lambda - \lambda_1)^{\alpha_1} \cdot (\lambda - \lambda_2)^{\alpha_2} \cdot \dots \cdot (\lambda - \lambda_k)^{\alpha_k}$$

$\underbrace{\hspace{1cm}}$ distinct e-values

$\underbrace{\hspace{1cm}}$ with multiplicities

$\underbrace{\hspace{1cm}}$

$\underbrace{\hspace{1cm}}$

$\underbrace{\hspace{1cm}}$

$$\dim N(A - \lambda_j I) = \alpha_j \text{ for all } j$$

$$\Rightarrow A = \left(\begin{array}{c|c|c}
\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_1 \end{pmatrix} & \begin{pmatrix} & & \\ & \ddots & \\ & & \lambda_2 \end{pmatrix} & \dots \\
\hline
N(A - \lambda_1 I) & N(A - \lambda_2 I) & \dots
\end{array} \right) \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_2 & \\ & & & \ddots & \lambda_{k-1} \\ & & & & \lambda_k \end{pmatrix} U'$$

$$\Rightarrow A = \left(\underbrace{\begin{pmatrix} \text{for basis } N(A-\lambda_1 I) \\ \vdots \\ \text{for basis } N(A-\lambda_n I) \end{pmatrix}}_{U} \cdots \right) \left(\begin{matrix} \text{times} \\ \dim N(A-\lambda_1 I) \\ \vdots \\ \dim N(A-\lambda_n I) \end{matrix} \right) U'$$

Corollary: If an $n \times n$ matrix A has n distinct eigenvalues, then A must be diagonalizable.

But not every diagonalizable matrix has n distinct eigenvalues.

eg, $I = \begin{pmatrix} 1 & & 0 \\ 0 & \ddots & \\ & & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

e-value 1 , multiplicity n $\lambda_1=1$, multiplicity $\alpha_1=3$
 $\lambda_2=2$, $\alpha_2=3$

Even if a matrix can be diagonalized, its eigenvectors might not be orthogonal.

Eg, $\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$ E-value 2 E-vector $(0, 1)$ ↗ not orthogonal!
 1
 $N\left(\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = N\left(\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}\right) = \text{Span}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$

Why doesn't Gram-Schmidt help?

- Performed on eigenvectors with different eigenvalues, it will output orthogonal vectors spanning the same space, but they won't (in general) still be eigenvectors.

E.g, $\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ $\xrightarrow{\text{Gram-Schmidt}}$ $\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$
 not an eigenvector of $\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$!

Example:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \neq$$

$$A A^T = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

but A is diagonalizable!

$$\lambda_1 = 1$$

$$\lambda_2 = 0$$

$$\text{Case } (0, 0) \quad \lambda_2 = 0$$

(e.g., since $\text{Trace}(A) = 1 = \lambda_1 + \lambda_2$
and $\text{Det}(A) = 0 = \lambda_1 \cdot \lambda_2$)

All distinct eigenvalues \Rightarrow diagonalizable.

Today: Lots of matrices are diagonalizable, with orthonormal eigenvectors.

For example, all symmetric matrices ($A = A^T$).
(And real symmetric matrices even have real eigenvalues)

Recall: Adjoint = conjugate transpose

$$\begin{pmatrix} a+bi & c+di \\ e+fi & g+hi \end{pmatrix}^* = \begin{pmatrix} a-bi & e-fi \\ c-di & g-hi \end{pmatrix}$$

(same as transpose for real matrices).

THEOREM: A has a complete, orthogonal set of eigenvectors

$$\Updownarrow A^T A = AA^T.$$

(Definition: A is "normal" $\Leftrightarrow A^T A = AA^T$.)

Proof:

↓: One direction is trivial. Assume A has a complete, orthonormal set of eigenvectors. Letting

$$U = \left(\left| \begin{array}{c} \text{set of} \\ \text{ortho-e-vectors} \\ \hline \end{array} \right| \right),$$

$$A = U D U^{-1},$$

where D is a diagonal matrix of the eigenvalues.

Since U is unitary, $U^{-1} = U^*$. Hence

$$AA^T = (UDU^*)(UDU^*)^*$$

$$\begin{aligned}
 &= UDU^+UD^+U^+ \\
 &= UDD^+U^+ \\
 A^+A &= (UDU^+)^+(UDU^+) \\
 &= UD^+U^+UDU^+ \\
 &= UDD^+U^+
 \end{aligned}$$

These are equal since $DD^+ = D^+D$ both just have the squared magnitudes $|\lambda_i|^2$ along the diagonal.
 (Every diagonal matrix is normal.) ✓

The other direction (A normal $\Rightarrow A = UDU^-$) is much more interesting. First let me prove two claims:

Claim 1: A normal $\Rightarrow R(A) = R(A^T)$.

(Note: If $R(A) \neq R(A^T)$, then there is no hope of finding a basis of eigenvectors.

A maps rowspace, $R(A^T)$, to columnspace, $R(A)$.
 If these spaces are different, then A does more than just scale some vectors.

Proof:

We have seen already that

$$N(A) = N(A^+A)$$

and applying this to A^+ gives

$$N(A^+) = N(AA^+)$$

(since $(A^+)^+ = A$). Hence

$$\begin{aligned}
 R(A^T) &= N(A)^+ = N(A^+A)^+ = N(AA^+)^+ = N(A^+)^+ = R(A) \\
 &\text{rank-nullity} \quad \text{normal} \quad \text{rank-nullity}
 \end{aligned}$$

✓ □

Claim 2: A normal \Rightarrow there exists a unitary matrix U such that

$$U^+AU = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}$$

for some nonsingular matrix C

Proof: By Claim 1, $R(A) = R(A^+)$, and $N(A) = N(A^+) = R(A)^{\perp}$.
 Let $U = \begin{pmatrix} \text{basis for } R(A) & | & \text{basis for } N(A)^{\perp} \\ \hline \end{pmatrix}$.

Since its columns are orthonormal, U is unitary: $U^+ = U^{-1}$.

$$\begin{aligned} U^+ A U &= \begin{pmatrix} \text{basis for } R(A) \\ \hline \text{basis for } N(A^+) \end{pmatrix} A \begin{pmatrix} \text{basis for } R(A) & | & \text{basis for } N(A)^{\perp} \\ \hline \end{pmatrix} \\ &= \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \quad \text{since } A \cdot \begin{pmatrix} \text{vector in } R(A^+) \\ \hline \end{pmatrix} = \begin{pmatrix} \text{vector in } R(A) \\ \hline \end{pmatrix} \perp N(A^+) \\ &\quad \text{and } A \cdot \begin{pmatrix} \text{vector in } N(A) \\ \hline \end{pmatrix} = 0 \\ C &= \underbrace{\begin{pmatrix} \text{basis for } R(A) \end{pmatrix}}_{\text{unitary}} A \begin{pmatrix} \text{basis for } R(A) \\ \hline \end{pmatrix} \end{aligned}$$

□

Now we're ready to prove the interesting direction:

$$\begin{array}{ll} \text{Theorem: } A^+ A = A A^+ & \text{Claim 2: } A^+ A = A A^+ \\ \Rightarrow A = U D U^+ & \text{using } \Rightarrow U^+ A U = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}_{\text{nonsingular}} \\ \text{unitary diagonal} & \end{array}$$

Proof:

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be A 's distinct eigenvalues.

$$\begin{aligned} A \text{ normal} &\Rightarrow A - \lambda_1 I \text{ normal} \\ ((A - \lambda_1 I)(A - \lambda_1 I)^+) &= (A - \lambda_1 I)^+(A - \lambda_1 I)) \\ \xrightarrow{\text{Claim 2}} \exists U_1 \text{ s.t.} & \\ U_1^+ (A - \lambda_1 I) U_1 &= \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} \\ \Rightarrow U_1^+ A U_1 &= \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} + \lambda_1 U_1^+ \underline{I} \\ &= \begin{pmatrix} C_1 + \lambda_1 I & 0 \\ 0 & \lambda_1 I \end{pmatrix} \end{aligned}$$

Let $A_1 = C_1 + \lambda_1 I$.

- Observe:
- λ_1 is not an eigenvalue of A_1 (or $A_1 - \lambda_1 I$ would be singular)
 - $\lambda_2, \dots, \lambda_k$ are still eigenvalues of A_1 ,
 since conjugating A by U_1 does
 not change the set of e-values, and
 $\lambda_2, \dots, \lambda_k$ are definitely not e-values of
 the second block $\lambda_1 I$
 - A_1 is normal!
 b/c $U_1^* A_1 U_1 = \begin{pmatrix} A_1 & 0 \\ 0 & \lambda_1 I \end{pmatrix}$ is normal
 —conjugating by a unitary does not change normality
 Since the matrix is block-diagonal, each block must
 be normal separately.

Therefore we can just recurse: find a unitary U_2 so

$$U_2^* (A_1 - \lambda_2 I) U_2 = \begin{pmatrix} C_2 & 0 \\ \underbrace{0}_{\text{nonsingular}} & 0 \end{pmatrix}$$

$$\Rightarrow U_2^* A_1 U_2 = \begin{pmatrix} C_2 + \lambda_2 I & 0 \\ 0 & \lambda_2 I \end{pmatrix}$$

A_3 ! etc.

Putting everything together, we find that for

$$U^+ = \dots \begin{pmatrix} U_3^* & 0 \\ 0 & I \\ 0 & I \end{pmatrix} \begin{pmatrix} U_2^* & 0 \\ 0 & I \end{pmatrix} U_1^+,$$

↑
to leave
the $\lambda_1 I$ term
unchanged

$$U^* A U = \begin{pmatrix} \lambda_k I & & & 0 \\ & \lambda_{k-1} I & \cdots & \\ & & \ddots & \lambda_2 I \\ 0 & & & \lambda_1 I \end{pmatrix}, \quad \square$$

Important: The theorem ($A^* A = A A^* \Rightarrow$ unitarily diagonalizable)
 is very important. So is the proof technique: Find one
 eigenspace, split it off, and recurse with the remainder.

Example: How can we use the power method to find the second-largest magnitude eigenvalue and the corresponding eigenvector?

One approach, in Matlab:

```
% using the power method to find the second-largest-magnitude eigenvalue
eigenvector
n = 100;
A = randn(n,n);
A = A + A';           % symmetric matrix => normal matrix

% first find the principal eigenvector using the power method
x = randn(n,1);
for j = 1:10000
    x = A * x;
    x = x / norm(x);
    x';
end
A*x ./ x      % using component-wise division, check that we've found an e-vector

% this starts with a vector perpendicular to the principal eigenvector,
% but numerical errors explode, causing it to be pushed parallel to the
% principal eigenvector
y = randn(n,1);
y = y - (x'*y)*x;
for j = 1:10000
    y = A * y;
    y = y / norm(y);
    y';
end
A * y ./ y

% to get the power method to work, we need to project orthogonal to the
% principal eigenvector after every step (or at least occasionally)
y = randn(n,1);
y = y - (x'*y)*x;
for j = 1:10000
    y = A * y;
    y = y - (x'*y)*x;
    y = y / norm(y);
    y';
end
A * y ./ y

% we can also find the k largest-magnitude eigenvalues simultaneously, using
the Gram-Schmidt procedure at every step of the power method
```

THEOREM: A has a complete, orthogonal set of eigenvectors

$$A^\dagger A \Updownarrow AA^\dagger. \quad ("A \text{ is normal}")$$

$$\Rightarrow A = U D U^\dagger$$

↑ ↓
 diagonal w/ e-values
 unitary w/
 e-vector columns

Examples:

- Every diagonal matrix is normal ($U=I$ above)
- No upper- or lower-triangular matrix is normal, unless it is diagonal! (see the homework)

```
>> n = 4;
>> A = randn(n,n);
>> A = A + A'; % make the matrix symmetric (and hence normal)
>> [U, D] = eigs(A)
```

$$U = \begin{matrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{matrix}$$

0.6207	-0.5958	0.1753	-0.4785
0.6988	0.1282	-0.1866	0.6786
-0.1355	-0.3777	-0.9151	-0.0408
-0.3287	-0.6971	0.3116	0.5558

the e-vectors are orthonormal:

>> $U' * U$

ans =

$$D = \begin{matrix} \lambda_1 & & & \\ & \lambda_2 & 0 & \\ & 0 & 3.7712 & 0 \\ & 0 & 0 & \lambda_3 \\ & 0 & 0 & 0 \end{matrix}$$

-7.2760	0	0	0
0	3.7712	0	0
0	0	1.8551	0
0	0	0	-0.3472

$$\begin{matrix} 1.0000 & 0 & 0.0000 & 0.0000 \\ 0 & 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & -0.0000 \\ 0.0000 & 0.0000 & -0.0000 & 1.0000 \end{matrix}$$

>> $A * U - U * D$

ans =

$$1.0e-14 *$$

0.3553	-0.1776	0.0056	-0.1832
-0.0888	-0.0722	0.1554	-0.3608
-0.1221	-0.0444	0.1110	0.0073
0.0444	0	-0.0888	-0.0416

$$A\vec{v}_j = \lambda_j \vec{v}_j$$

Proposition: A normal $\Rightarrow \|A\| = \max_{\text{eigenvalues } \lambda} |\lambda|$.

Proof:

Recall $\|A\| = \max_{x: \|x\|=1} \|Ax\|$. Since unitaries don't change lengths,

$$\begin{aligned}\|A\| &= \|UDU^{-1}\| \\ &= \|D\| \\ &= \max_i |D_{i,i}| \quad \square\end{aligned}$$

This proposition is false for non-normal matrices,
e.g., both eigenvalues of $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ are 0,
but $\left\| \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\| = 1 \neq 0$.

Key point:

Normal matrix \Rightarrow Different eigenspaces
are orthogonal

Because the nullspace (e-value 0 e-space)
satisfies

$$N(A) = N(A^T) = R(A)^\perp$$

and all other e-spaces
are in $R(A)$