Matrix Perturbation

**Part One**

Thin: Let $X$ ind. columns

\[
X = \begin{bmatrix}
\end{bmatrix}
\]

$R = R(X)$

$x$ is $A$-invariant iff

$Y^H A X = 0$

where $Y$'s columns span $\mathbb{R}$

iff: $x$ $A$-invariant

$A x \in \mathbb{R}$

$A^T x \perp x$

$R(A X) \perp R(Y)$

$Y^H A X = 0$
It is $A$-invariant.

$X_1$ be orthonormal for $X$

$(X_1, Y_2)$ be unitary (so columns of $(X_1, Y_2)$ are orthonormal for entire space)

Then,

$$(X_1, Y_2)^T A (X_1, Y_2) = \begin{bmatrix} L_1 & H \\ 0 & L_2 \end{bmatrix}$$

$H$ some matrix

$$AX_1 = X_1 L_1$$

$L_1$ is a representation of $A$ on $X = \text{span}(X_1)$, w.r.t. the basis $X_1$

Defn: $X$ is a simple invariant subspace

1. Invariant (w.r.t. $A$)
2. $\mathcal{L}(L_1) \cap \mathcal{L}(L_2) = \emptyset$

$\lambda$-value of $L_1$
Matrix Perturbation

Given: \( A, B \) (for \( L_1, L_2 \))

define: \( T_{A,B}(X) = AX - X \cdot B \)

Then: \( T \) is non-singular if

\[ \mathcal{L}(A) \cap \mathcal{L}(B) = \emptyset \]

[In class, proved one direction]

Corollary: \( \mathcal{L}(T) = \mathcal{L}(A) - \mathcal{L}(B) \)

Exercise: \( T \) is a linear operator.

How would we represent \( T \) as multiplication by a matrix? 

\[ \Rightarrow \text{Any linear operator in finite dimension can be represented as multiplication by a matrix.} \]
Spectral Resolution

Then: If $\mathbf{x}$ is $A$-invariant and simple, then $\mathbf{x}^\perp$ is $A$-invariant.

Exercise: Find an example of an invariant subspace $\mathbf{F}$ where $\mathbf{F}^\perp$ is not invariant.

In particular, if similarity transformation $T$ may not be unitary.

\[ (x_1, x_2)^{-1} A (x_1, x_2) = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix} \]

\[ \exists \mathbf{x}_2, \mathbf{y}_1 \quad \text{st.} \]

\[ (x_1, x_2)^{-1} = (y_1, y_2)^{\dagger} \]

Define: $A = x_1 L_1 y_1^\dagger + x_2 L_2 y_2^\dagger$
Matrix Perturbation

\[ P_i = X_i Y_i^H \quad X_i, Y_i \text{ play} \]
\[ P_2 = X_2 Y_2^H \quad \text{the role of left and right e-vectors} \]

Exercise: What happens in 1-dim case, i.e., check what happens when \( \mathbb{F} \) is a simple invariant 1-dim subspace.

Exercise: \( P_i^2 = P_i \quad i = 1, 2 \)
\[ P_1 P_2 = P_2 P_1 = 0 \]
\[ A = P_1 A P_1 + P_2 A P_2 \]

\( P_i \) is called the spectral projection of \( \mathbb{F} \).

We will see in PS exercise, that \( \| P_i \| \) controls sensitivity of eigenvalues in \( \mathbb{F} \).

Will look at 1-d case.
Matrix Permutation

Last Time

1. Simple invariant subspace

2. \( T : \mathbf{X} \rightarrow A \mathbf{X} - \mathbf{X} \mathbf{B} \)

\[
(x_1, x_2)^T A (x_1, x_2) = \begin{pmatrix} L_1 \ H \\ \ 0 \ L_2 \end{pmatrix}
\]

"How close are \( L_1 \) and \( L_2 \)"

\( \mathcal{M} = \mathcal{R}(x_1) \) is called simple if \( \mathcal{Z}(L_1) \cap \mathcal{Z}(L_2) = \emptyset \)

\( T_{L_1, L_2} (\mathbf{x}) = L_2 \mathbf{x} - \mathbf{x} L_1 \)

if \( T \) is non-singular

Then \( \mathcal{Z}(L_1) \cap \mathcal{Z}(L_2) = \emptyset \)

and conversely.
Approximation Problems

Let $X_i$ be orthonormal for an approximate $A$-invariant subspace.

$$L = X_i^H A X_i$$

Recall: If $X = R(X_i)$ were $A$-invariant,

then, $A X_i = X_i L_i$

$$R = A X_i - X_i L$$

so $X_i$ is not (quite)

$A$-invariant.

Q: How close is $R(X_i) = X$

to being $A$-invariant, as a function of $||R||$
Matrix Perturbation

\[ X_1 \text{ on } \mathcal{X} \]

\[(X_1, Y_2)^\top A (X_1, Y_2) = \begin{pmatrix} L_1 & H \\ G & L_2 \end{pmatrix} \]

Recall: if \( G = 0 \iff \mathcal{X} = R(X_1) \) is \( A \)-invariant,

\[ G = Y_2^\top A X_1 \]

Want to find \( A \)-invariant subspace "near" \( \mathcal{X} \).

\[ \hat{X}_1 = (X_1 + Y_2 P)(I + P^\top P)^{-1/2} \]

\[ \hat{Y}_2 = (Y_2 - X_1 P^\top)(I + P^\top P)^{-1/2} \]

Exercise: \( \begin{pmatrix} \hat{X}_1 \\ \hat{Y}_2 \end{pmatrix} \) unitary (inherits from \( \begin{pmatrix} X_1 \\ Y_2 \end{pmatrix} \))

Want to choose \( P \) so that \( R(\hat{X}_1) \) is \( A \)-invariant.

\( R(\hat{X}_1) \) is \( A \)-invariant \( \iff \)

\[ \hat{Y}_2^\top A \hat{X}_1 = 0 \]
\[ Y^* A X_i = 0 \iff \]

\[ P L_1 - L_2 P = G + PHP = 0 \]

\[ \iff P L_1 - L_2 P = G - PHP \]

\[ \tilde{T}_{L_1, L_2} (P) = G - PHP \]

\( P \) makes \( R(X_i) \) \( A \)-invariant exactly when it solves the non-linear equation:

\[ \tilde{T}(P) = G - PHP \]
Theorem: Assume that $\mathcal{X}(L_1) \cap \mathcal{X}(L_2) = \phi$ implies $T$ is non-singular.

[Want conditions on magnitude of $H, G$ that tell us how far we are from invariant subspace.]

Norms used here are "consistent."

\[
\gamma = \|G\|, \\
i_1 = \|H\|, \\
\delta = \text{sep}(L_1, L_2) = \inf_{\|Q\|=1} \|T(Q)\| > 0
\]

I. If $\frac{\gamma i_1}{\delta^2} < \frac{1}{4}$ then \exists $P$ s.t.

\[\hat{\mathcal{X}}_1 \text{ is } A \text{-invariant} \]

II. And, $\|P\| \leq 2 \frac{\gamma}{\delta}$

in fact: $\|P\| \leq \frac{2\gamma}{\delta + \sqrt{\delta^2 - 4\gamma i_1}} \leq 2 \frac{\gamma}{\delta}$

III. $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2$ span for $A$-invariant simple

\[
\hat{L}_1 = (I + \Phi^H \Phi)^{1/2} (L_1 + HP) (I + \Phi^H \Phi)^{-1/2} \\
\hat{L}_2 = (I + \Phi^H \Phi)^{1/2} (L_2 - HP) (I + \Phi^H \Phi)^{-1/2}
\]
Matrix Perturbation

Recap + How to use

\[(X_1, Y_2)^T A (X_1, Y_2) = \begin{pmatrix} L_1 & H \\ G_r & L_2 \end{pmatrix}\]

Thm says: If \( \frac{\delta^2}{\delta^2} < \frac{1}{4} \), \( \exists \ P \), \( \| P \| \leq 2 \frac{\delta}{\delta} \sum \)

\[s.t. \quad \hat{X}_1 = (X_1 + Y_2 P) (I + P^T P)^{-1/2}, \quad \hat{Y}_2 = (Y_2 - X_1 P^T) (I + P^T P)^{-1/2}\]

\( R(\hat{X}_1) \) is \( A \)-invariant

Q: How close is \( \hat{X} \Rightarrow R(\hat{X}_1) \) to \( R(X_1) \)?

Distance btw subspaces: Canonical Angles

\[Y_2^T \hat{X}_1 \rightarrow \text{singular values, if this matrix are the sines of the Canonical Angles.}\]

\[Y_2^T \hat{X}_1 = P (I + P^T P)^{-1/2}\]

If \( P \) has singular values: \( \tilde{\pi}_1, \tilde{\pi}_2, \ldots \)

then

\[\sin \Theta_i = \frac{\tilde{\pi}_i}{\sqrt{1 + \tilde{\pi}_i^2}} \leq \tilde{\pi}_i\]

Thm tells us:

Canonical angles btw \( R(X_1) \), \( R(\hat{X}_1) \), \( \| P \| \leq 2 \frac{\delta}{\delta} \sum \)
Matrix Perturbation

\[ P_1 = \text{projection onto } R(X_1) \]
\[ \hat{P}_1 = \text{onto } R(\hat{X}_1) \]

\[ \| P_1 - \hat{P}_1 \| \]

Fact: \[ \| P_1 - \hat{P}_1 \|_2 = \sin \theta_1 \leq \| P \|_2 \]

\[ \text{operator norm} \]

If \[ \gamma = \| G \|_2, \quad \eta = \| H \|_2 \]
\[ \delta = \text{sep}(L_1, L_2) = \inf_{\| Q \|_2 = 1} \| T(Q) \|_2 \]

Then \[ \| P \|_2 \leq 2 \frac{\delta}{\gamma} \]

\[ \Rightarrow \| P_1 - \hat{P}_1 \|_2 \leq 2 \frac{\delta}{\gamma} \]
Matrix Perturbation

What's left:

1. Proof of the Theorem

2. How to apply the approximation to matrix Perturbation.

3. What happens if we do things simply when $A$ is Hermitian.
PART THREE

Matrix Perturbation

1. Perturbation Theory for General Matrices
2. Case of Hermitian Matrices

A + \xi have to check that \xi is in an invariant subspace.

Easier, but we do not need to check if \xi is in a subspace.

Theorem: A matrix

\[(X_1, y_2) \text{ in } R(X_1) \text{ is } A\text{-invariant.}\]

\[(X_1, y_2)^{\dagger} A (X_1, y_2) = \begin{pmatrix} L_1 & H \\ 0 & L_2 \end{pmatrix}\]

Invariance \(\iff\) \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.

Consider \(A + \xi\),

\[(X_1, y_2)^{\dagger} (X_1, y_2) = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}\]

Does \(A + \xi\) have an invariant subspace that is close to \(\xi = R(X_1)\), and how close?
\[(X, Y_2)^H (A + E) (X, Y_2) = \begin{bmatrix} L_1 + E_{11} & (H + E_{12}) \\ E_{12} & L_2 + E_{22} \end{bmatrix} \]

From previous (approx.) theorem:

Let \[\|L - H\| = \|L_2 - H_2\|\]

\[\|X\| = \|E_{21}\|\]

\[\|Y\| = \|H + E_{11}\| \leq \|H\| \|1\| + \|E_{11}\|\]

\[\delta = \text{sep}(L_1 + E_{11}, L_2 + E_{22})\]

This expression is not ideal. The theorem then promises that

\[\frac{\delta}{\sqrt{\delta^2 + \gamma^2}} \leq \frac{\gamma}{\sqrt{\delta^2 + \gamma^2}}\]

If \[\frac{\delta}{\sqrt{\delta^2 + \gamma^2}} < \frac{\gamma}{\sqrt{\delta^2 + \gamma^2}}\]

Then \(A\) is invertible

\[\frac{\delta}{\sqrt{\delta^2 + \gamma^2}}\]

In particular, \(\exists P, \forall \|P\|_2 \leq 2\frac{\delta}{\gamma}\)

As we saw when dividing how to use the approximate theorem,

\[\tan \Theta_{\text{max}} \leq \|P\|_2 = \frac{\gamma}{\sqrt{\delta^2 + \gamma^2}}\]

pf: immediate.
Issue: Results we just gave depend on
\[
\delta = \text{sep}(L_1 + E_n, L_2 + E_2)
\]
and ideally we want \( \delta \) that depends
on \( A(L_1, L_2) \), and on size
of \( E \). \( ||E||_1, ||E||_2 \).

Need to better understand: \[\text{sep}\]

We need the following continuity
for \( \text{sep} \):

Thus:

\[
\frac{\text{sep}(L, M) - ||E||}{- ||E||} \leq \text{sep}(L + E, M + E) \leq \text{sep}(L, M) + ||E|| + ||E||
\]

pf: Exercise.
Thm: \((X, Y_2)^T A (X, Y_2) = \begin{pmatrix} L_1 & H \\ 0 & L_2 \end{pmatrix}\)

If \(E\) perturbation \(\xi\)

\((X, Y_2)^T E (X, Y_2) = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}\)

Then writing: 
\[
\tilde{y} = \|E_{21}\|
\]

\[
\tilde{\gamma} = \|H\| + \|E_{12}\|
\]

\[
\tilde{\delta} = \text{sep}(L_1, L_2) - (\|E_{11}\| + \|E_{22}\|)
\]

Then, if \(\frac{\tilde{\gamma}^2}{\tilde{\delta}^2} < \frac{1}{4}\), \(E\) \((4 + \varepsilon)\)-inv.

such \(\delta\) with that

\[
\tan \theta_{\max} \leq 2 \frac{\tilde{\gamma}}{\tilde{\delta}}
\]

Summary: Perturbation results depend

on:

1. Size of perturbation
2. Matrix \(A\) invariant subspace \(K\)

through \(T_{L_1, L_2}: X \rightarrow L_1 X - X L_2\)

\[
\inf \|T(Q)\| \leq \text{sep}(L_1, L_2)
\]

\(\|Q\| = 1\)
Thm: For any square matrices $L_1, L_2$

$$\inf_{||T|| = 1} ||T|| = \sup \{|\lambda(L_1) - \Lambda(L_2)|\}$$

If $\text{sep}(L_1, L_2) = 0$

$$\iff \exists \text{ singular}$$

$$\iff \lambda(L_1) \cap \lambda(L_2) \neq \emptyset$$

If $\text{sep}(L_1, L_2) > 0 \Rightarrow \exists \text{ non-singular}$

Exercise: $\text{sep}^{-1}(L_1, L_2) = \sup_{||T|| = 1} ||T^{-1}||$

Exercise: For any consistent norm,

$$p(M) \leq ||M||$$

$$\Rightarrow p(T^{-1}) \leq ||T^{-1}||$$

Also have shown: $\lambda(T) = \lambda(L_1) - \lambda(L_2)$

$$\Rightarrow \text{sep}^{-1}(L_1, L_2) = ||T^{-1}|| \geq p(T^{-1}) = \max|\lambda(L_1) - \lambda(L_2)|$$

This is the statement of the Thm.
Says: If $L_1, L_2$ have $\epsilon$-values that are close

$\Rightarrow \text{sep}(L_1, L_2)$ is small

$\Rightarrow \frac{1}{\epsilon}$ big

$\Rightarrow$ Perturbation bound large

For Hermitian matrices: the condition that $\epsilon$-values of $L_1, L_2$ for $\epsilon$-perturb is necessary and sufficient to guarantee stability.

Exercise: For non-H, it's possible for $\text{sep}(L_1, L_2)$ to be small but to have $\epsilon$-values of $L_1, L_2$ well-separated.

Compute $\text{sep}(L_1, L_2) \leq \min \{ \|L(L_1) - 7L_2\| \}$ for $L_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $L_2 = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$. 
Part Four

Perturbation results for Hermitian matrices

Easier: \( M \) is Hermitian, then there is a o.n.b. \( \mathbf{f} \) e-vectors

\[ M = A, \quad M = A + E \]
\[ \Rightarrow \text{Always have invariant subspaces.} \]

So can focus on computing bases—no longer need extra conditions to guarantee existence of invariant subspaces.

“Direct Bounds” \( \Rightarrow (\text{Davis} \& \text{Kahan}) \)

**Theorem:** If \( L, M \) Hermitian, then

\[
\text{sep}_F(L, M) = \inf_{Q\mathbf{f}^{-1}} \| T(Q) \|_F
\]

\[
= \min \| Q(L) - Q(M) \|_F
\]
Thm: \( A \) Hermitian,

\[
(X_1 X_2)^T A (X_1 X_2) = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}
\]

unitary

\( Z = R(X_1) \) is \( A \)-invariant.

Suppose \( Z \) same dim as \( X_1 \),

o.n. columns

How close is \( R(Z) \) to \( R(X_1) \)?

Let \( M \) be any k\( \times \)k matrix

\( M \) is an approximate rep of \( A \)

\[ Z \]

Recall: \( X_1 \) is \( A \)-invariant

\[
AX_1 = X_1 L \quad \text{rep of } A \text{ on } X_1, \ldots, X_k \text{ bases}
\]

\[
(AX_1 - X_1 L = 0)
\]

Approximate: \( AZ - ZM = R \)

Q: How close is \( R(X_1) \), \( R(Z) \)?
Matrix Perturbation

Note: This is like defining a perturbation \( E = \begin{pmatrix} M - Z_1 & 0 \\ 0 & 0 \end{pmatrix} \)

\[ (X, X_2)^T A (X, X_2) = \begin{pmatrix} l_1 & 0 \\ 0 & l_2 \end{pmatrix} \]

\[ AZ - ZM = R \]

\( A \) and \( Z \) have the same dimension as \( X \).

Let \( \delta = \min \left| \lambda(L_2) - \lambda(M) \right| > 0 \)

Then:
\[ \| \sin \theta [R(X_1), R(Z)] \|_F \leq \frac{\| R \|_F}{\delta} \]

diagonal matrix with canonical

singular values \( R(X_1), R(Z) \) on diagonal
Matrix Perturbation

Let \( \begin{pmatrix} x_1 & x_2 \end{pmatrix}^T A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)

Proof: \( R = A z - z M \)

\[ \Rightarrow X_z^H R = X_z^H A z - X_z^H z M \]

Using \( S = L_z X_z^H Z - X_z^H z M \)

Hence:

\[ S = L_z (X_z^H Z) - (X_z^H z) M \]

\[ = \begin{pmatrix} T_{L_z M} \end{pmatrix} \begin{pmatrix} X_z^H Z \end{pmatrix} \]

Now: \( S = \text{sep}_F (L_z, M) \)

\[ \Rightarrow \lim_{\|Q\|_F \rightarrow 1} \|T(Q)\|_F = S \cdot \|Q\|_F \]

\( Q = X_z^H z \)

\[ \delta \|X_z^H z\|_F \leq \|T(Q)\|_F \]

\[ \Rightarrow \|X_z^H z\|_F \leq \frac{\|R\|_F}{\delta} \]

\( R(x_1), R(z) \)

\[ \|X_z^H z\|_F = \| \sin \Theta [R(x_1), R(z)] \|_F \leq \frac{\|R\|_F}{\delta} \]
$$R = AZ - ZM$$

Then:
$$\| \sin \Theta [R(x_1), R(x_2)] \|_F \leq \frac{\| R \|_F}{\delta}$$

$$\delta = \text{sep}_F = \min \| x^*_2 - x^*_1 \|_F$$

For op-norm bound

When is this useful? This assumption is tough.

If dim of $X_1$ (or $X_2$) is small, this result can give useful bounds.

But, if dim of $X_1$ is large, then

the Frobenius norm becomes large

If dim $X_1 = 1$ then Frob. norm = $\| \cdot \|_F$ norm.

For large dimensional invariant subspaces:

would like results in terms of operator norm not Frobenius norm.

For this: we need further restrictions on exactly how the evectors of $L_2, M$ are separated.
Lemma: Let $\| \cdot \|$ be a consistent norm, with

\[
\| A \| \leq \alpha
\]
\[
\| B^{-1} \|^{-1} \geq \alpha + \delta \quad \text{for } \delta > 0
\]

If $Ax - xB = C$

Then $\| x \| \leq \frac{\| C \|}{\delta}$

pf: By consistency: $\| Ax \| \leq \| A \| \cdot \| x \| = \alpha \| x \|

\| xB \| \geq (\alpha + \delta) \| x \|

\Rightarrow C = Ax - xB

\| C \| \geq \| Bx \| - \| A \| \| x \| \geq \delta \| x \|

How will we use this:

In prov. thm: $\hat{R} = AZ - 2M$

Here: need to translate the Lemma's assumption on $\| A \|, \| B^{-1} \|^{-1}$ into e-value condition.
Matrix Perturbation

Theorem: In the same context as in the previous sin-Θ theorem,

If \( \lambda_1(\mathbf{M}) \leq [\alpha, \beta] \) and for \( \delta > 0 \)

\[
\lambda(L_2) = \mathbb{R} \setminus [\alpha - \delta, \beta + \delta]
\]

Then, for any unitarily invariant norm \( \| \|_{\text{op}}, \| \|_2, \| \|_F \)

\[
\| \sin \Theta [R(x_1), R(z)] \| \leq \frac{\| R \|}{\delta}
\]

If \( \| R \| = \| R \|_2 \) then this theorem says

\[
\sin \Theta_{\max} \leq \frac{\| R \|}{\delta}
\]
pf: Check that WLOG we can assume that $\alpha = -\beta$

and $X(M) \subseteq [-\alpha, \alpha]$

$X(L_2) \subseteq (-\infty, -\alpha-\delta) \cup (\alpha+\delta, \infty)$

To apply previous lemma:

$||M|| \leq \alpha, \quad ||L_2|| \geq \alpha + \delta$

Similarly to proof of previous sin Theta theorem,

since

$\begin{pmatrix} X_1 & X_2 \end{pmatrix}^H A \begin{pmatrix} X_1 & X_2 \end{pmatrix} = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}$

and $R = A Z - Z M$

$\Rightarrow X_2^H R = X_2^H A Z - X_2^H Z M$

$= L_2 (X_2^H Z) - (X_2^H Z) M$

Now we can apply the lemma above using the operator norm:

$||X_2^H Z|| \leq \frac{||X_2^H R||}{\delta} = \frac{||R||}{\delta}$

In other words:

$||\sin \Theta [R(X_1), R(Z)]|| \leq \frac{||R||}{\delta}$
Matrix Perturbation

Last two results: stated in terms of any \( Z, M \).

Let's apply above \( \sin \Theta \) then directly to perturbation:

**Thm:** \( M, M + \Delta \) Hermitian

hence we can write:

\[
M = (x^*_1, x^*_2)^T \begin{bmatrix} M_0 & \hat{M}_0^* \\ \hat{M}_0 & M_1 \end{bmatrix} (x_1, x_2)
\]

\[
M + \Delta = (y^*_1, y^*_2)^T \begin{bmatrix} \hat{M}_0 & \hat{M}_0^* \\ \hat{M}_0^* & M_1 \end{bmatrix} (y_1, y_2)
\]

where \( ||\Delta|| \) is given

\[
\mathbb{K}(M_0) \subseteq [a, b], \quad \mathbb{K}(M_0^*) \subseteq (-\infty, a+\delta) \cup (b+\delta, \infty)
\]

\[
\mathbb{K} = R(x_1), \quad M - \text{invariant}
\]

\[
\mathbb{Y} = R(y_1), \quad (M+\Delta) - \text{invariant}
\]

**Let** \( R = M y_1 - y_1 \hat{M}_0 \)

\[
||R|| \leq ||M y_1 - y_1 \hat{M}_0|| = ||(M+\Delta) y_1 - y_1 M_0 - \Delta y_1|| \\
\leq ||(M+\Delta) y_1 - y_1 M_0|| + ||\Delta y_1|| = 0 + ||\Delta||
\]

\[
\Rightarrow || \sin \Theta [R(x_1), R(y_1)] || \leq \frac{||\Delta||}{\delta}
\]

PS is immediate.