Schur Transformation, General Case

Schur basis is NOT UNIQUE. But, sum of squares of FF weights = Tr $WW^\dagger - \sum_i |\lambda_i|^2$ is a unitary invariant. One natural scalar measure is this, or this over $Tr WW^\dagger$, or squareroot of either.

More generally: in Schur basis $W - \Lambda = N$ is upper triangular element; $\min_{\text{Schur decompositions}} |N| =$ Henrici’s “departure from normality” $\text{dep}(W)$; $|N|$ same for all decompositions for Frobenious norm. Another measure: $|WW^\dagger - W^\dagger W|$.

Trefethan & Embree:

In the end, all scalar measures of nonnormality suffer from a basic limitation: Nonnormality is too complex to be summarized in a single number. Even when a matrix is ‘highly nonnormal’, one must know the geometry of that nonnormality in order to judge whether it will have important consequences for a given application."

For $J$ random IID matrix, elements variance $1/N$: $Tr JJ^\dagger = N^2 \frac{1}{N} = N$. Sum over eigenvalues

$$= N \times \left\langle |\lambda|^2 \right\rangle$$

$$= N \int_0^1 drr^3 \int d\theta$$

$$= N^2/2$$

Example: submatrices simultaneously diagonalizable by unitary transform

e.g. if translation invariant: $\rightarrow 2 \times 2$ submatrices for each freq.

Example: Rajan and Abbott

Example: $\begin{pmatrix} W_E & -W_I \\ W_E & -W_I \end{pmatrix}$ go to sum and diff basis of those eigs; sum modes have eigs $\lambda_i^E - \lambda_i^I$; also $N$ zero eigs; diff modes give $\lambda_i^E + \lambda_i^I$ times sum modes.

SLIDE

Schur Basis, 2 × 2 case

Assume two distinct eigenvectors, $e_1, e_2$; complete $e_1$ with Schur basis vector $q$:

$$q = \frac{e_2 - (e_1 \cdot e_2)e_1}{\sqrt{1 - |e_1 \cdot e_2|^2}}$$
Then compute
\[ Wq = \frac{\lambda_2 e_2 - \lambda_1 (e_1 \cdot e_2) e_1}{\sqrt{1 - |e_1 \cdot e_2|^2}} = \lambda_2 q + \frac{(\lambda_2 - \lambda_1) (e_1 \cdot e_2)}{\sqrt{1 - |e_1 \cdot e_2|^2}} e_1 \] (5)

Thus, in the Schur basis \((e_1, q)\), \(W\) takes the upper triangular form
\[ W = \begin{pmatrix} \lambda_1 & \beta \\ 0 & \lambda^+ \end{pmatrix} \]
with
\[ \beta = \frac{(\lambda_2 - \lambda_1) (e_1 \cdot e_2)}{\sqrt{1 - |e_1 \cdot e_2|^2}} \] (6)

\(\beta\) is
- Large when \(|e_1 \cdot e_2|\) is close to one, i.e. when the angle between the eigenvectors is small;
- Zero when the matrix is normal, so that \(|e_1 \cdot e_2| = 0\).

(It also becomes zero if \(\lambda_1 = \lambda_2\), but this means that the matrix is normal, assuming there are two distinct eigenvectors.)

For biological weight matrix: for both eigenvectors, real parts of both elements are same sign: sum mode; \(q\) is difference mode\(^1\).

Compute \(\beta\) for weight matrix:
- Eigs real: \(x^2_+ - w_{EI}w_{IE} > 0\)
  \[ |\beta|^2 = \text{Tr} WW^\dagger - |\lambda_+|^2 - |\lambda_-|^2 \]
  \[ = w_{EE}^2 + w_{EI}^2 + w_{IE}^2 + w_{II}^2 - 2(x_+^2 + x_-^2 - w_{EI}w_{IE}) \]
  \[ = (w_{EI} + w_{IE})^2 \] (12)

Note \(\beta > 0\), therefore \(\beta = w_{EI} + w_{IE}\). With real eigs, \(\beta\) is large when \(W\) deviates strongly from Hermitian.

\(^1\)Eigenvalues:
\[ \lambda_\pm = \frac{w_{EE} - w_{II}}{2} \pm \sqrt{\left(\frac{w_{EE} + w_{II}}{2}\right)^2 - w_{EI}w_{IE}} \] (7)
\[ \equiv x_\pm \sqrt{x_+^2 - w_{EI}w_{IE}} \] (8)

Eigenvectors:
\[ e_\pm \propto \begin{pmatrix} 1 \\ (\lambda_+ + w_{II})/w_{EI} \end{pmatrix} = \begin{pmatrix} 1 \\ x_+ \pm \sqrt{x_+^2 - w_{EI}w_{IE}}/w_{EI} \end{pmatrix} \] (9)

Take \(|e_\pm| = 1\). Note: both elements have real parts > 0; sum mode. Assume eigenvectors are distinct.

Let Schur basis be \(\{e_-, q\}\). \(q \propto \begin{pmatrix} (\lambda_+^* + w_{II})/w_{EI} \\ -1 \end{pmatrix}\): difference mode.
• Eigs complex: \( w_{EI}w_{IE} - x_+^2 > 0 \)

\[
|\beta|^2 = \text{Tr} \, W W^\dagger - |\lambda_+|^2 - |\lambda_-|^2
\]

\[
= w_{EE}^2 + w_{EI}^2 + w_{IE}^2 + w_{II}^2 - 2(x_-^2 + w_{EI}w_{IE} - x_+^2)
\]

\[
= (w_{EI} - w_{IE})^2 + (w_{EE} + w_{II})^2
\]

(13)

(14)

(15)

With complex eigs, \( \beta \) is large when \( W \) deviates strongly from anti-Hermitian.

**Pseudospectra**

Spectra: \( \exists v \text{ s.t. } Wv = zv \) for scalar \( z \). In terms of resolvent of \( W \), \( R_W(z) = (W - z1)^{-1} \),

spectrum is \( \sigma(W) = \{ z : ||R_W(z)|| = \infty \} \)

\( \epsilon \)-Pseudospectra: \( \sigma_\epsilon(W) = \{ z : ||R_W(z)|| > \frac{1}{\epsilon} \} \), i.e. for vector-induced norm, \( \exists v \text{ s.t. } |Wv - zv| < \epsilon. \)

For vector-induced norm: equivalent definition is \( \sigma_\epsilon(W) = \{ z : \exists dW, ||dW|| < \epsilon \text{ s.t. } z \in \sigma(W + dW) \}. \) For normal matrix, pseudospectrum is just the set of \( \epsilon \)-balls around the spectrum. For non-normal, it always includes that but is generally larger, can be much larger – spectrum very non-robust to perturbation.

Example: \( M = \begin{pmatrix} \lambda_1 & \beta \\ 0 & \lambda_2 \end{pmatrix} \). Perturb with \( \begin{pmatrix} 0 & 0 \\ \epsilon & 0 \end{pmatrix} \), size \( \epsilon. \) Spectrum is perturbed from \( \{ \lambda_1, \lambda_2 \} \) to

\[
\frac{\lambda_1 + \lambda_2}{2} \pm \sqrt{\left(\frac{\lambda_1 - \lambda_2}{2}\right)^2 + \beta \epsilon}
\]

(16)

\[
= \frac{\lambda_1 + \lambda_2}{2} \pm \frac{\lambda_1 - \lambda_2}{2} \sqrt{1 + \frac{4\beta \epsilon}{(\lambda_1 - \lambda_2)^2}}
\]

(17)

\[
= \left\{ \begin{array}{l}
\frac{\lambda_1 + \lambda_2}{2} \left( \sqrt{1 + \frac{4\beta \epsilon}{(\lambda_1 - \lambda_2)^2}} - 1 \right), \\
\frac{\lambda_2 - \lambda_1}{2} \left( \sqrt{1 + \frac{4\beta \epsilon}{(\lambda_1 - \lambda_2)^2}} - 1 \right)
\end{array} \right\}
\]

(18)

(19)

\[
\rightarrow \beta \rightarrow \infty \quad \frac{\lambda_1 + \lambda_2}{2} \pm \sqrt{\frac{\beta}{\epsilon}}
\]

(20)

**Kreiss Bound, bounds on transients**

Pseudospectra give a lower bound on the size of transients of a dynamics \( \frac{dx}{dt} = Wx \), where all the eigenvalues of \( W \) have negative real part:

\[
\sup_{t \geq 0} ||e^{tW}|| \geq \alpha_\epsilon(W)/\epsilon \quad \text{for all } \epsilon > 0
\]

(21)
where \(\alpha_e(W)\) is the largest real part of an element of the \(\epsilon\)-pseudospectrum of \(W\). This leads directly to the Kreiss bound: defining

\[
\mathcal{K}(W) = \sup_{\epsilon > 0} \frac{\alpha_e(W)}{\epsilon} = \sup_{\Re z > 0} (\Re(z)) \|(z - W)^{-1}\|
\]  

then

\[
\sup_{t \geq 0} ||e^{tW}|| \geq \mathcal{K}(W)
\]  

also

\[
||e^{tW}|| \leq eN\mathcal{K}(W)
\]

**Other bounds:**

Numerical range (or field of values) \(W(W)\): the set \(\{x^\dagger W x \text{ for } |x| = 1\};\) equals eigenvalues of symmetric part of \(W\), \((W + W^\dagger)/2\). Contains convex hull of spectrum; = convex hull for normal matrix. Numerical abscissa \(\omega(W)\) = maximum real part of numerical range = max eig of \((W + W^\dagger)/2\).

\(\alpha(W)\): spectral abscissa: max real part of spectrum

For all \(t\), \(e^{t\alpha(W)} \leq ||e^{tW}|| \leq e^{t\omega(W)}\)

\[
||e^{tW}|| \leq e^{t\omega(W)} \quad \forall t
\]

\[
||e^{tW}|| \leq e^{t\omega(W)} + o(t) \text{ as } t \to 0
\]

**Proof of Kreiss Bound:**

The bound in Eq. 21 is established as follows. There is always some real \(\omega\) and \(M \geq 1\) such that \(||e^{tW}|| \leq Me^{\omega t}\) for all \(t \geq 0\). \(\omega\) can be any number larger than the largest real part of an eigenvalue of \(W\) (you may have to choose \(M\) large to compensate for transients at finite times), so when all eigenvalues have negative real part we can always choose \(\omega < 0\).

Then, for any \(z\) with \(\Re z > w\), and in particular for any \(z\) with \(\Re z > 0\),

\[
(z - W)^{-1} = \int_0^\infty dt e^{-t(z - W)} = \int_0^\infty dt e^{-tz} e^{-tW}
\]  

This implies

\[
||(z - W)^{-1}|| \leq \int_0^\infty dt |e^{-tz}| ||e^{-tW}||
\]  

Let \(S = \sup_{t \geq 0} ||e^{tW}||\). Then

\[
||(z - W)^{-1}|| \leq S\int_0^\infty dt |e^{-tz}| = S/(\Re(z))
\]  

or

\[
\sup_{t \geq 0} ||e^{tW}|| \geq (\Re(z))||(z - W)^{-1}||
\]
How long do pseudospectra act like spectra?

Pseudospectrum can “act like eigenvalues” during transient period, if move to positive real part by more than $\epsilon$. If $z$ is in $\sigma_\epsilon(W)$, $\Re(z) > 0$, $\Re(z)/\epsilon > 1$, then for any $\tau > 0$,

$$
\sup_{0 < t \leq \tau} \|e^{tW}\| \geq e^{\Re(z)\tau} / \left(1 + \epsilon \frac{e^{\Re(z)\tau}}{\Re(z)} - 1\right)
$$

(28)

Expression in paren is close to 1 for $e^{\Re(z)\tau} \ll 1 + \Re(z)/\epsilon$, i.e. $\tau < \frac{\log(1 + \Re(z)/\epsilon)}{\Re(z)}$. On those time scales, $z$ acts like an eigenvalue.
Bounds on pseudospectra

Eigenvalue condition number: \( \kappa(\lambda) = \frac{\|r_l\|}{\|l_r\|} = \frac{1}{\cos(\theta_{lr})} \) where \( l \) and \( r \) are left and right eigenvectors associated with eigenvalues. When all eigenvalues distinct: then as \( \epsilon \to 0 \)

\[
\sigma_\epsilon(W) \subseteq \bigcup_{j=1}^N \left( \lambda_j + \Delta_{\epsilon\kappa(\lambda_j)} + O(\epsilon^2) \right)
\]  

(37)

and for all \( \epsilon > 0 \), for vector-induced norms,

\[
\sigma_\epsilon(W) \subseteq \bigcup_{j=1}^N \left( \lambda_j + \Delta_{\epsilon\kappa(\lambda_j)} \right)
\]  

(38)

Pseudospectra and numerical range:

\[
\sigma_\epsilon(W) \subseteq W(W) + \Delta_\epsilon
\]

and

\[
W(W) = \text{intersection of all closed half-planes } H \in \text{complex plane satisfying}
\]

\[
\sigma_\epsilon(W) \subseteq H + \Delta_\epsilon \quad \forall \epsilon > 0.
\]

---

**Proof for \( \epsilon \to 0 \):** Consider \( W(t) = W + tE \) where \( \|E\| = 1 \). Perturbs to \( \lambda(t) \), \( r(t) \). Because \( \lambda \) is simple root of characteristic equation, can expand as Taylor series in \( t \). Note \( l \) is orthogonal to all other right eigenvectors, so writing \( r(t) = r + dE(t) \) with \( dE \) equal to \( t \) times a linear combination of the other right eigenvectors, then \( l \cdot r(t) = 1 \cdot r \). So then:

\[
1^l W(t)r(t) = \lambda(t)1^l r = 1^l Wr(t) + t1^l Er(t)
\]

(29)

\[
= \lambda l^1 r + t l^1 Er(t)
\]

(30)

So

\[
|\lambda - \lambda(t)| = \frac{|t| 1^l Er(t)|}{|l^1 r|}
\]

(31)

\[
\leq |t| \frac{|1^l |r(t)|}{|l^1 r|}
\]

(32)

\[
= |t| \frac{|1^l |r|}{|l^1 r|} + O(|t|^2)
\]

(33)

\[
= |t| \kappa(\lambda) + O(|t|^2)
\]

(34)

**Proof for general \( \epsilon \):** For each eig, normalize such that \( 1^l r = 1 \). Define \( P_j = r_j l_j^1 ; \) projection, \( P^2 = P \).

Note

\[
\|P\| = \frac{\|l^1 r\|}{\|l^1 r\|} = \frac{\|r\|_x}{\|l^1 r\|} = \frac{\|r\|_l}{\|l^1 r\|} = \kappa(\lambda)
\]

(35)

where chose \( x = 1/\|l\| \).

\( z \in \sigma_\epsilon(W) \) implies \( \exists x \) with \( |x| = 1 \), \( E \) with \( \|E\| < \epsilon \), s.t. \( (W + E)x = zx \). Then \( PEx = P(z - W)x = (z - \lambda)x \). Provided \( 1^l x \neq 0 \), this implies for any eigenvalue, eigenvector that

\[
|\lambda - z| = \frac{|PEx|}{Px} < \epsilon \frac{|P|}{|Px|}
\]

(36)

But \( 1 = |x| = |\sum_j P_j x| \leq \sum_j |P_j x| \), so there must be at least one \( j \) with \( |P_j x| > 1/N \). For this \( j \),

\[
|\lambda_j - z| < \epsilon N\|P_j\| = \epsilon N\kappa(\lambda_j)
\]

as required.