The Barycenter of the Numerical Range of an Operator

S. Allen Broughton - Rose-Hulman Institute of Technology
Roger Lautzenheiser - Rose-Hulman Institute of Technology
Thomas Werne - Jet Propulsion Laboratory

ISU Math and CS Research Seminar, November 28, 2007
There are three main parts of the talk

- Introduction, basic examples, and properties.
- Discussion of the 2-D case and the Toeplitz-Hausdorff compactness-convexity result.
- Discussion of the barycenter and proof of barycenter theorem.
Notation and definitions

- $V$ is a Hilbert space, but just really $\mathbb{C}^n$ for our purposes
- $X = (x_1, \ldots, x_n), \ Y = (y_1, \ldots, y_n) \in V$ are any two vectors
- and $\langle X, Y \rangle = x_1\overline{y}_1 + \cdots + x_n\overline{y}_n$ is the standard Hermitian scalar product of $X$ and $Y$
- if $Y^* = $ conjugate transpose, then $\langle X, Y \rangle = Y^*X$ for column vectors
- $\|X\| = \sqrt{\langle X, X \rangle}$
- $B_n = B(V) = \{X \in V : \|X\| \leq 1\}$ is the unit ball in $V$
- $\partial B_n = \partial B(V) = \{X \in V : \|X\| = 1\}$ is the unit sphere in $V$
A : \mathbb{V} \rightarrow \mathbb{V} is any operator, but really just an \( n \times n \) matrix

\( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( A \).

Recall the equation for the spectrum average

\[
\frac{1}{n} \sum_{i=1}^{n} \lambda_i = \frac{1}{n} \text{trace}(A)
\]

also define the map

\[ f_A : \partial B_n \rightarrow \mathbb{C}, \text{ by } f_A(X) = \langle AX, X \rangle. \]
Definition

Let $A : V \rightarrow V$ be a bounded linear operator of the Hilbert space $V$. The numerical range $W(A)$ is the subset in the complex plane defined by

$$W(A) = \{ \langle AX, X \rangle : \|X\| = 1 \}$$
what does $W(A)$ look like?
simple properties

**Proposition**

- For a finite dimensional space $V$ the numerical range $W(A)$ is a compact subset of the plane.
- The numerical range $W(A)$ contains the eigenvalues of $A$.

- The numerical range $W(A)$ is the continuous image of $\partial B_n$ under $f_A$.
- Let $AX = \lambda X$ for some $\lambda$ and some unit vector $X$. Then $\langle AX, X \rangle = \langle \lambda X, X \rangle = \lambda \langle X, X \rangle = \lambda$. 
**Proposition**

*If A is a diagonal matrix then W(A) is the convex hull of the set of eigenvalues.*

**Proof sketch**

- Assume that $A = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and that $X = (x_1, \ldots, x_n)$.
- Then $\langle AX, X \rangle = \sum_{i=1}^{n} ||x_i||^2 \lambda_i$
- As $\sum_{i=1}^{n} ||x_i||^2 = 1$ then $\langle AX, X \rangle$ is a convex linear combination of the eigenvalues.
The following properties are useful in $W(A)$ calculations.

**Proposition**

- If $U$ is a unitary matrix then $W(UAU^{-1}) = W(A)$.
- For complex constants $a, b$, $W(ai + bA) = a + bW(A)$.

For unitary $U$, $U^{-1} = U^*$. Setting $Y = UX$ we get

\[
\langle UA^{-1}Y, Y \rangle = \langle UA^{-1}UX, UX \rangle = \langle UAX, UX \rangle = \langle AX, X \rangle.
\]

As $X$ varies completely over the sphere so does $Y = UX$.

\[
\langle (ai + bA)X, X \rangle = a\langle X, X \rangle + b\langle AX, X \rangle = a + b\langle AX, X \rangle
\]
Restricting to a subspace, is a useful computational technique. Here is specific computational formulation.

**Proposition**

Let $W \subseteq V$ be subspace and let $X_1, \ldots, X_m$ be an orthonormal basis of $W$. Let $B$ be the $m \times m$ matrix defined by

$$B_{i,j} = \langle AX_i, X_j \rangle$$

Then

$$W(B) \subseteq W(A).$$
First properties and examples

**restriction to a subspace - 2**

**Proof sketch**

- Set $P = [X_1 \ X_2 \ \cdots \ X_m]$, then by orthogonality $P^*P = I_m$.
- Let $Z \in \partial B_m$ and $X = PZ = \sum_{i=1}^{m} z_iX_i$.
- Then $||X|| = 1$ as $\langle X, X \rangle = X^*X = Z^*P^*PZ = Z^*Z = 1$.
- $X = PZ$ defines an isometry from $\partial B_m$ to $W \cap \partial B_n$.
- For $X \in W \cap \partial B_n$, $\langle AX, X \rangle = \langle APZ, PZ \rangle = \langle (P^*AP)Z, Z \rangle$.
- $W(P^*AP) = \{\langle AX, X \rangle : X \in W \cap \partial B_n\} \subseteq W(A)$
- The $i, j$ entry of $P^*AP$ is $X_i^*AX_j = \langle AX_i, X_j \rangle = B_{i,j}$
2D case

**Proposition**

*If A is a $2 \times 2$ matrix then $W(A)$ is a filled ellipse with the eigenvalues at the foci.*

We give a proof sketch since it uses basic techniques used studying the numerical range.
proof sketch -1

Proof sketch

- Select unitary $U$ such that $UAU^{-1}$ is upper triangular - use Schur’s Lemma. So we assume that $A$ is upper triangular.

- Let $\tau = \text{trace}(A)/2$. Then there is a unit complex scalar $\upsilon$ such that $\upsilon(A - \tau I)$ has eigenvalues $\pm a$ for real $a$. Thus, for some complex $b$, $A$ has the form

$$A = \begin{bmatrix} a & 2b \\ 0 & -a \end{bmatrix}.$$ 

- The effect of the above transformation is a rigid motion in the plane, taking ellipses to ellipses, foci to foci and eigenvalues to eigenvalues.
2D case

proof sketch - 2

Next, use the unitary similarity

\[
\begin{bmatrix}
e^{i\phi} & 0 \\
0 & e^{i\psi}
\end{bmatrix}
\begin{bmatrix}
a & b \\
0 & -a
\end{bmatrix}
\begin{bmatrix}
e^{-i\phi} & 0 \\
0 & e^{-i\psi}
\end{bmatrix} =
\begin{bmatrix}
a & e^{i(\phi-\psi)}b \\
0 & -a
\end{bmatrix}
\]

so that we may assume that \(b\) is real non-negative.

A typical unit vector \(X\) in \(\mathbb{C}^2\) has the form

\[
X = \begin{bmatrix}
\cos(\theta)e^{i\phi} \\
\sin(\theta)e^{i\psi}
\end{bmatrix}
\]

and so

\[
\langle AX, X \rangle = \begin{bmatrix}
\cos \theta e^{-i\phi} & \sin \theta e^{-i\psi}
\end{bmatrix}
\begin{bmatrix}
a & b \\
0 & -a
\end{bmatrix}
\begin{bmatrix}
\cos(\theta)e^{i\phi} \\
\sin(\theta)e^{i\psi}
\end{bmatrix}
\]

or

\[
\langle AX, X \rangle = a(\cos^2 \theta - \sin^2 \theta) + 2b \cos \theta \sin \theta e^{i(\psi-\phi)}
\]
or for suitable $\alpha, \beta$

$$\langle AX, X \rangle = a \cos(\alpha) + b \sin(\alpha)e^{i\beta}$$

$$\langle AX, X \rangle = a \cos(\alpha) + b \sin(\alpha)\cos(\beta) + ib \sin(\alpha)\sin(\beta)$$

With some work, one can show that as $\alpha, \beta$ vary the ellipse

$$\frac{x^2}{a^2 + b^2} + \frac{y^2}{b^2} \leq 1$$

is swept out.

The foci of this ellipse are at $-a$ and $a$, the eigenvalues of $A$. 
The Toeplitz-Hausdorff theorem dramatically reduces the possibilities for the shape of the numerical range of a matrix.

**Theorem**

*The numerical range of $W(A)$ of a matrix $A$ is a compact, convex subset of the plane.*
proof sketch

- Let $X$ and $Y$ be two vectors such that $\langle AX, X \rangle$ and $\langle AY, Y \rangle$ are distinct.
- Let $W \subseteq V$ be the linear span of $X$ and $Y$ and let $X_1, X_2$ be an orthonormal basis of $W$.
- By previous proposition, the set of values $\langle AZ, Z \rangle$ for all unit vectors $Z$ in $W$ is the same as the numerical range $W(B)$ of the $2 \times 2$ matrix

\[
B = \begin{bmatrix}
\langle AX_1, X_1 \rangle & \langle AX_1, X_2 \rangle \\
\langle AX_2, X_1 \rangle & \langle AX_2, X_2 \rangle
\end{bmatrix}
\]

- Thus $\langle AX, X \rangle$ and $\langle AY, Y \rangle$ are contained in an ellipse contained in $W(A)$. 
The average of the eigenvalues appear to be at the center of $W(A)$.

Proven to be true for the 2 $\times$ 2 case.
How to generate pictures

- Select a large number of vectors $X_1, X_2, \ldots, X_N$ uniformly distributed on $\partial B_n$
- Plot $\langle AX_i, X_i \rangle$ for $N$ different vectors. Here are two examples.
some observations

- Points are not uniformly distributed on $W(A)$, so the standard centroid is not the right idea for the “center” of $W(A)$.

- The sample average $\frac{1}{N} \sum_{i=1}^{N} \langle AX_i, X_i \rangle$ seems to be very close to the spectrum average $\frac{1}{n} \sum_{i=1}^{n} \lambda_i$.

- The result above appears to hold true even if the vectors are only distributed “symmetrically”.
Definition

We define the barycenter (center of mass) of $W(A)$ to be

$$BW(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \langle AX_i, X_i \rangle$$

where the $X_i$'s are chosen from the uniform distribution on the boundary of the unit ball in $\mathbb{C}^n$
The $X_i$’s are uniformly distributed $\partial B_n$ if or each closed subset $U$ of $\partial B_n$,

$$\lim_{N \to \infty} \frac{\#\{i : X_i \in U\}}{N} = \frac{\text{vol}(U)}{\text{vol}(\partial B_n)},$$

where $\text{vol}(U)$ is the volume of $U$ computed as a subset of the $\partial B_n$.
We get an integral definition

\[ BW(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \langle AX_i, X_i \rangle = \int_{\partial B_n} \langle AX, X \rangle \, d\omega. \]

Define this planar density on \( W(A) \)

\[ \delta(z) = \lim_{r \to 0} \frac{\omega(f_A^{-1}(\Delta_r(z)))}{\pi r^2} \]

with \( \Delta_r(z) = \{ w \in \mathbb{C} : \| w - z \| \leq r \} \).

Then \( BW(A) \) has a planar integral definition

\[ BW(A) = \int_{\partial B_n} \langle AX, X \rangle \, d\omega = \int_{W(A)} z \delta(z) \, dx \, dy \]
The following theorem characterizes the barycenter.

**Theorem**

The barycenter $BW(A)$ of the numerical range $W(A)$ is given by:

$$BW(A) = \frac{\text{tr}(A)}{n} = \frac{1}{n} \sum_{i=1}^{n} \lambda_i.$$
proof sketch -1

- From the definitions.

\[ BW(A) = \int_{\partial B_n} \langle AX, X \rangle \, d\omega = \sum_{i,j} \int_{\partial B_n} a_{i,j} x_i \bar{x}_j \, d\omega \]

- We need only prove

\[ \int_{\partial B_n} x_i \bar{x}_j \, d\omega = \frac{1}{n} \delta_{i,j} \]
Proof sketch - 2

Now some setup

- Define the functions

\[ f_i(X) = x_i \bar{x}_i, \quad f_{i,j}(X) = x_i \bar{x}_j \]

- Note that

\[ \sum_i f_i(X) = \sum_i x_i \bar{x}_i = \langle X, X \rangle = 1 \]

- Also define unitary operators (transpositions and symmetries along coordinate axes)

\[ U_{i,j} : (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) \longrightarrow (x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n) \]

for any distinct \( i, j \) and

\[ V_i : (x_1, \ldots, x_i, \ldots, x_n) \longrightarrow (x_1, \ldots, -x_i, \ldots, x_n) \]
From invariance

\[ \int_{\partial B_n} x_i x_i \, d\omega = \int_{\partial B_n} f_i(X) \, d\omega = \int_{\partial B_n} f_i(U_{i,j}X) \, d\omega = \int_{\partial B_n} f_j(X) \, d\omega = \int_{\partial B_n} x_j x_j \, d\omega \]

and so

\[ n \int_{\partial B_n} x_i x_i \, d\mu = \int_{\partial B_n} \sum_j x_j x_j \, d\mu = \int_{\partial B_n} 1 \, d\mu = 1 \]

proving

\[ \int_{\partial B_n} x_i x_i \, d\omega = \frac{1}{n} \]
Now assuming $i \neq j$,

$$\int_{\partial B_n} x_i \overline{x}_j d\omega = \int_{\partial B_n} f_{i,j}(X) d\omega = \int_{\partial B_n} f_{i,j}(V_i X) d\omega$$

$$\int_{\partial B_n} f_{i,j}(V_i X) d\omega = \int_{\partial B_n} -f_{i,j}(X) d\omega = -\int_{\partial B_n} x_i \overline{x}_j d\omega.$$ 

and hence

$$\int_{\partial B_n} x_i \overline{x}_j d\omega = 0.$$
Remark

If the vectors are randomly chosen from any probability distribution \( \mu \) on the sphere invariant under the \( V_i \) and \( U_{i,j} \) then

\[
BW(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \langle AX_i, X_i \rangle = \int_{\partial B_n} \langle AX, X \rangle \, d\mu
\]
Thank you.
Any questions?