Finite-dimensional spectral theory

part II: understanding the spectrum (and singular values)

Ed Bueler

MATH 617 Functional Analysis

Spring 2020

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

what happened in part I

- see part I first: bueler.github.io/M617S20/slides1.pdf
- *definition.* for a square matrix $A \in \mathbb{C}^{n\times n}$, the *spectrum* is the set

$$\sigma(A) = \left\{\lambda \in \mathbb{C} \,\middle|\, Av = \lambda v \text{ for some } v \neq 0\right\}$$

- we proved:

  $A = QTQ^*$ *Schur decomposition* for any $A \in \mathbb{C}^{n\times n}$
  $A = Q\Lambda Q^*$ *spectral theorem* for normal ($AA^* = A^*A$) matrices

  where Q is unitary, T is upper-triangular, and Λ is diagonal
  - both decompositions "reveal" the spectrum:

  $$\sigma(A) = \{\text{diagonal entries of } T \text{ or } \Lambda\}$$

  - spectral theorem for hermitian matrices is sometimes called the *principal axis decomposition* for quadratic forms

goal

extend the spectral theorem to ∞-dimensions

- only possible for linear operators on Hilbert spaces H
  - inner product needed for adjoints and unitaries
  - unitary maps needed because they preserve vector space *and* metric *and* adjoint structures
- textbook (Muscat) extends to compact normal operators on H
  - the spectrum is eigenvalues (almost exclusively)
- recommended text (B. Hall, *Quantum Theory for Mathematicians*) extends further to bounded (continuous) normal operators on H
  - spectrum is not only eigenvalues
  - statement of theorem uses projector-valued measures
- Hall also extends to unbounded normal operators on H
  - but we won't get there ...
- the Schur decomposition has no straightforward extension

important class: unitary matrices

- back to matrices!

Definition

$U \in \mathbb{C}^{n\times n}$ is *unitary* if $U^* U = I$

Lemma

Consider $\mathbb{C}^n$ as a inner product space with $\langle v, w\rangle = v^* w$ and $\|v\|_2 = \sqrt{\langle v, v\rangle}$. Suppose U is linear map on $\mathbb{C}^n$. The following are equivalent:

- *U is unitary*
- *expressed in the standard basis, the columns of U are ON basis of $\mathbb{C}^n$*
- *$\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v \in \mathbb{C}^n$*
- *$\|Uv\|_2 = \|v\|_2$ for all $v \in \mathbb{C}^n$*
- *U is a metric-space isometry*

important class: normal matrices

Definition

$A \in \mathbb{C}^{n\times n}$ is *normal* if $A^* A = AA^*$

- includes: hermitian ($A^* = A$), unitary, skew-hermitian ($A^* = -A$)

Lemma

Consider $\mathbb{C}^n$ as a inner product space with $\langle v, w\rangle = v^* w$ and $\|v\|_2 = \sqrt{\langle v, v\rangle}$. Suppose A is linear map on $\mathbb{C}^n$. The following are equivalent:

- *A is normal*
- *$\|Ax\|_2 = \|A^* x\|_2$ for all x*
- *exists an ON basis of eigenvectors of A*
- *exists Q unitary and Λ diagonal so that $A = Q\Lambda Q^*$ (spectral theorem)*

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

power series of matrices

- suppose A is diagonalizable: $A = S\Lambda S^{-1}$
    - where S is invertible and Λ is diagonal
    - diagonal entries of Λ are eigenvalues of A
    - if A is normal (e.g. hermitian) then choose $S = Q$ unitary so $S^{-1} = Q^*$
- powers of A:

$$A^k = S\Lambda S^{-1} S\Lambda S^{-1} S\Lambda S^{-1} \cdots S\Lambda S^{-1} = S\Lambda^k S^{-1}$$

- if $f(z)$ is a power series then we can create $f(A)$:

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \qquad \implies \qquad f(A) = \sum_{n=0}^{\infty} c_n A^n = S\left(\sum_{n=0}^{\infty} c_n \Lambda^n\right) S^{-1}$$

$$= S \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} S^{-1}$$

- for example: $$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = S \begin{bmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_n} \end{bmatrix} S^{-1}$$

what does “functional calculus” mean?

- given $A \in \mathbb{C}^{n\times n}$, a (finite-dimensional) *functional calculus* is algebraic-structure-preserving map from a set of functions $f(z)$ defined on $\mathbb{C}$ to matrices $f(A) \in \mathbb{C}^{n\times n}$
- example: for $f(z)$ analytic,

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n \qquad \implies \qquad f(A) = \sum_{n=0}^{\infty} c_n (A - z_0 I)^n$$

$$= S \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} S^{-1}$$

- but ...
  - does the matrix power series $f(A) = \sum_{n=0}^{\infty} c_n (A - z_0 I)^n$ converge? **reasonable question**
  - does $f(z)$ have to be analytic anyway? **no**

norms of powers

- for any induced norm:

$$\|A^k\| \le \|A\|^k$$

- if A is diagonalizable then in any induced norm

$$\|A^k\| = \|S\Lambda^k S^{-1}\| \le \kappa(S) \max_{\lambda\in\sigma(A)} |\lambda|^k = \kappa(S)\rho(A)^k$$

  - $\kappa(S) = \|S\|\|S^{-1}\|$ is the *condition number* of S
  - $\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|$ is the *spectral radius* of A
  - $\rho(A) \le \|A\|$
- *corollary.* if A is diagonalizable and $\rho(A) < 1$ then $A^k \to 0$ as $k \to \infty$
  - actually this holds for all square A ... use the Schur or Jordan-canonical-form decompositions
- if A is normal then, because unitaries preserve 2-norm,

$$\|A^k\|_2 = \|Q\Lambda^k Q^*\|_2 = \max_{\lambda\in\sigma(A)} |\lambda|^k = \rho(A)^k$$

  - thus $\|A^k\|_2 = \|A\|_2^k$
  - note $\kappa_2(Q) = 1$ for a unitary matrix Q

convergence when $f(z)$ is analytic

does it converge? $$f(A) \overset{*}{=} \sum_{n=0}^{\infty} c_n (A - z_0 I)^n$$

Lemma

Suppose $f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n$ has radius of convergence $R > 0$. If $\|A - z_0 I\| < R$ in some induced norm then sum $*$ converges in that norm.

- if A is normal then $A = Q\Lambda Q^*$ so

$$\|A - z_0 I\|_2 = \max_{\lambda \in \sigma(A)} |\lambda - z_0| = \rho(A - z_0 I)$$

- in general $\rho(A - z_0 I) \leq \|A - z_0 I\|$ can be strict inequality

defining $f(z)$

- compare two ways of defining $f(A)$:

$$f(A) \overset{(1)}{=} \sum_{n=0}^{\infty} c_n (A - z_0 I)^n \qquad \text{and} \qquad f(A) \overset{(2)}{=} S \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} S^{-1}$$

- for (1) f needs to be analytic and have sufficiently-large radius of convergence relative to norm $\|A - z_0 I\|$
- for formula (2), A needs to be diagonalizable, but $f(z)$ does not need to be analytic ... it only needs to be defined on $\sigma(A)$

Theorem

If $A \in \mathbb{C}^{n\times n}$ is normal, if $\sigma(A) \subseteq \Omega \subseteq \mathbb{C}$, and if $f : \Omega \to \mathbb{C}$, then there is a unique matrix $f(A) \in \mathbb{C}^n$ so that:

1. *$f(A)$ is normal*
2. *$f(A)$ commutes with A*
3. *if $Av = \lambda v$ then $f(A)v = f(\lambda)v$*
4. $\|f(A)\|_2 = \max_{\lambda\in\sigma(A)} |f(\lambda)|$

proof. By the spectral theorem there is a unitary matrix Q and a diagonal matrix Λ so that $A = Q\Lambda A^*$, with columns of Q which are eigenvectors of A and all eigenvalues of A listed on the diagonal of Λ. Define

$$f(A) = Q \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} Q^*.$$

It has the stated properties. It is a unique because its action on a basis (eigenvectors of A) is determined by property 3.

the meaning of the functional calculus

- if A is normal then you can apply any function $f(z)$ to it, giving $f(A)$, as though A is "just like a complex number"
  - f merely has to be defined[1] on the finite set $\sigma(A)$
  - the matrix 2-norm behaves well: $\|f(A)\|_2 = \max_{\lambda\in\sigma(A)} |f(\lambda)|$
  - eigendecomposition is therefore powerful when A is normal!
- if A is diagonalizable then $f(A)$ can be *defined* the same:

$$f(A) = S \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} S^{-1}$$

  but surprising behavior is possible: $\|f(A)\| \gg \max_{\lambda\in\sigma(A)} |f(\lambda)|$
- if A is defective then what? revert to using power series just to define $f(A)$?

[1] In ∞-dimensions f needs some regularity. Thus there are separate wikipedia pages on *holomorphic functional calculus*, *continuous functional calculus*, and *borel functional calculus*.

functional calculus applications

1. suppose A is hermitian and we want to build a unitary matrix from it
   - A is normal and $\sigma(A) \subset \mathbb{R}$

solution 1. $f(z) = e^{iz}$ maps $\mathbb{R}$ to the unit circle so

$$U = e^{iA} \quad \text{is unitary}$$

solution 2. $f(z) = \dfrac{z+i}{z-i}$ maps $\mathbb{R}$ to the unit circle so

$$U = (A + iI)(A - iI)^{-1} \quad \text{is unitary}$$

2. suppose U is unitary and we want to build a hermitian matrix from it
   - U is normal and $\sigma(U) \subset S^1 = \{z \in \mathbb{C} \,:\, |z| = 1\}$

solution. $f(z) = \mathrm{Log}(z)$ maps the unit circle S^1 to the real line, so

$$A = \frac{1}{i}\,\mathrm{Log}(U) = -i\,\mathrm{Log}(U) \quad \text{is hermitian}$$

functional calculus applications: linear ODEs

3. given $A \in \mathbb{C}^{n\times n}$ normal, and given $y_0 \in \mathbb{C}$, solve
$$\frac{dy}{dt} = Ay, \qquad y(t_0) = y_0$$
for $y(t) \in \mathbb{C}^n$ on $t \in [t_0, t_f]$

solution. $y(t) = e^{tz}$ solves $dy/dt = zy$ so, using the functional calculus with $f(z) = e^{(t-t_0)z}$,
$$\begin{aligned} y(t) &= e^{(t-t_0)A} y_0 \\ &= \texttt{expm((t-t0)*A)*y0}, \\ \|y(t)\|_2 &= e^{(t-t_0)\omega(A)} \|y_0\|_2 \end{aligned}$$
where $\omega(A) = \max_{\lambda\in\sigma(A)} \operatorname{Re}\lambda$

- if A is diagonalizable $A = S\Lambda S^{-1}$ then the same applies ... except the norm of the solution includes $\kappa(S)$
- if A is defective then the general solution of the ODE system is *not* exponential

4. ∞-dimensional version: Schrödinger's equation in quantum mechanics

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

Definition

given $A \in \mathbb{C}^{n\times n}$ then $\mathbb{C} \setminus \sigma(A)$ is the *resolvent set*, and if $z \in \mathbb{C} \setminus \sigma(A)$ then

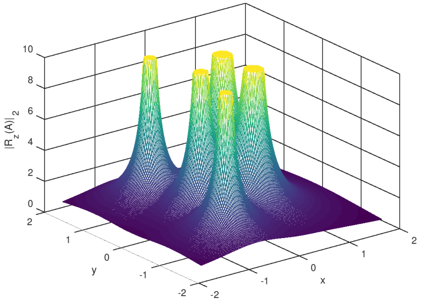
$$R_z(A) = (A - zI)^{-1}$$

is the *resolvent matrix*

- recall: $z \in \sigma(A)$ if and only if $A - zI$ is not invertible
- the resolvent set $\mathbb{C} \setminus \sigma(A)$ is open
- $R_0(A) = A^{-1}$ if $0 \notin \sigma(A)$
- $R_z(A)$ "resolves" the equation $Av - zv = b$

resolvent norms

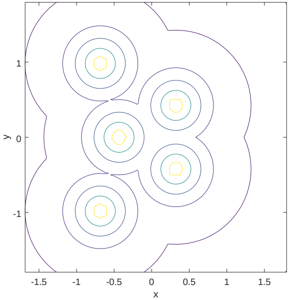
- if $A = S\Lambda S^{-1}$ is diagonalizable and $z \in \mathbb{C} \setminus \sigma(A)$ then
$$R_z(A) = \left(S\Lambda S^{-1} - zSIS^{-1}\right)^{-1} = S\left(\Lambda - zI\right)^{-1} S^{-1}$$
so in any induced norm
$$\|R_z(A)\| = \|S\|\|S^{-1}\|\|\left(\Lambda - zI\right)^{-1}\| = \kappa(S) \max_{\lambda\in\sigma(A)} |\lambda - z|^{-1}$$
- if A is normal then we can choose $S = Q$ unitary with $\kappa_2(Q) = 1$ so
$$\|R_z(A)\|_2 = \max_{\lambda\in\sigma(A)} |\lambda - z|^{-1}$$
- one may plot $g(z) = \|R_z(A)\|$

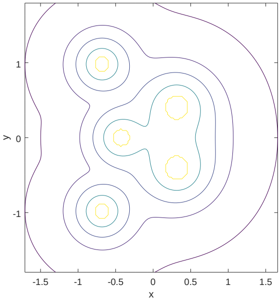


resolvent norms illustrated

- contours of $z \mapsto \|R_z(A)\|_2 = \|(A - zI)^{-1}\|_2$ is best spectral picture?

```
>> [A,B] = gennormal(5);   % A,B have same eigs; A normal but B not
>> resolveshow(A)          % normal case      (LEFT)
>> resolveshow(B)          % nonnormal case   (RIGHT)
```



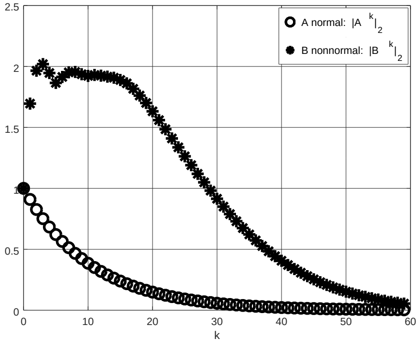


- last slide already proved contours would be round for normal A
- $\sigma_\epsilon(A) = \left\{z \in \mathbb{C} \,:\, \|(A - zI)^{-1}\|_2 \geq \epsilon^{-1}\right\}$ is the ϵ-*pseudospectrum* of A

nonnormal matrices, a warning

- facts and definitions:
  - $\|A^k\| \le \|A\|^k$ in any induced norm
  - $\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|$
  - if A is normal then $\|A^k\|_2 = (\|A\|_2)^k = \rho(A)^k$
  - if $\rho(A) < 1$ then $A^k \to 0$ as $k \to \infty$ *proof?*
- but if A is not normal and $\rho(A) < 1$ then $\|A^k\|_2$ *can be big* for a while
  - e.g. random 100×100 matrices A,B with $\rho(A) = \rho(B) < 1$

```
>> max(abs(eig(A)))
ans =  0.90909
>> max(abs(eig(B)))
ans =  0.90909
```



redefining "spectrum": nonexistence of resolvent

Definition

given $A \in \mathbb{C}^{n\times n}$, the *spectrum of A* is the set

$$\sigma(A) = \left\{\lambda \in \mathbb{C} \,\middle|\, A - \lambda I \text{ does not have a bounded inverse}\right\}$$

- in $\mathbb{C}^n$ this is the same as our original definition:

$$\sigma(A) = \left\{\lambda \in \mathbb{C} \,\middle|\, Av = \lambda v \text{ for some } v \neq 0\right\}$$

- in ∞-dimensions it is different because there exist one-to-one bounded operators which do not have bounded inverses
  - *example 1*: the one-to-one right-shift operator R on ℓ^1 has spectrum[2] $\sigma(R) = \{z \in \mathbb{C} : |z| \leq 1\}$, but it has no eigenvalues
  - *example 2*: the hermitian multiplication operator $(Mf)(x) = xf(x)$ on $L^2[0,1]$ has no eigenvalues but $\sigma(M) = [0,1]$

[2] we will prove this by showing that $\sigma(L) = \{z \in \mathbb{C} : |z| \leq 1\}$ for the left-shift operator $L = R^*$, based on eigenvalues, and that $\sigma(A^*) = \sigma(A)$ in a Banach algebra

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

orthogonal projectors

Definition

$P \in \mathbb{C}^{n\times n}$ is an *orthogonal projector* if $P^2 = P$ and $P^* = P$

- as for any projector ($P^2 = P$):

$$\ker P = \operatorname{im}(I - P), \quad \operatorname{im} P = \ker(I - P), \quad \mathbb{C}^n = \ker P \oplus \operatorname{im} P, \quad \sigma(P) \subset \{0, 1\}$$

- but for orthogonal projectors:

$$\ker P \perp \operatorname{im} P$$

  - *proof.* if $u \in \ker P$ and $v = Pz \in \operatorname{im} P$ then $u^* v = u^*(Pz) = (Pu)^* z = 0$

- orthogonal projectors are hermitian, thus normal
- examples:

$$0, \quad I, \quad P = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix}$$

constructing orthogonal projectors from ON vectors

- since P is hermitian and $\sigma(P) \subset \{0, 1\}$, the spectral theorem plus re-ordering of the columns of Q gives

$$P = Q\Lambda Q^* = Q \begin{bmatrix} \hat{I} & \\ & 0 \end{bmatrix} Q^* = \hat{Q}\hat{Q}^*$$

where $\hat{I}$ is a $k \times k$ identity and $\hat{Q}$ is a $n \times k$ matrix of columns of Q

Lemma

$P \in \mathbb{C}^{n\times n}$ is an orthogonal projector if and only if there exist ON vectors $q_1, \dots, q_k$, for $0 \le k \le n$, so that

$$P = \hat{Q}\hat{Q}^* \quad \text{and} \quad \hat{Q} = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \dots & q_k \end{array} \right] \in \mathbb{C}^{n\times k}$$

- hard direction of proof is above; easy direction: $(\hat{Q}\hat{Q}^*)^2 = \dots$
- note $\hat{Q}^*\hat{Q} = \hat{I}$
- rank 1 case: $\quad P = qq^* = (aa^*)/(a^*a)$
- construction from full-column-rank A: $\quad P = A(A^*A)^{-1}A^*$

spectral theorem = decomposition into projectors

- consider this calculation for A normal:

$$A = Q\Lambda Q^* = Q\left(\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}\right) Q^*$$

$$= Q\left(\begin{bmatrix} \lambda_1 & \\ & \end{bmatrix} + \dots + \begin{bmatrix} & \\ & \lambda_n \end{bmatrix}\right) Q^* = q_1 \lambda_1 q_1^* + \dots + q_n \lambda_n q_n^*$$

$$= \sum_{j=1}^{n} \lambda_j q_j q_j^*$$

  - A decomposes into a linear combination of rank-one orthogonal projectors

- thus normal matrices act on vectors like this:

$$Av = \sum_{j=1}^{n} \lambda_j q_j q_j^* v = \sum_{j=1}^{n} \lambda_j \langle q_j, v \rangle q_j$$

  - this formula appears in most applications of normal operators

resolution of the identity

- if A is normal then $A = \sum_{i=1}^{n} \lambda_i q_i q_i^*$ where $\{q_i\}$ are ON
- if A is normal then we can use its eigenvectors to decompose the identity:

$$I = QQ^* = \sum_{i=1}^{n} q_i q_i^*$$

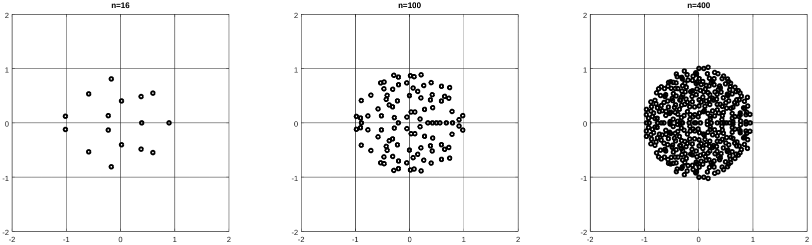
  - called a *resolution of the identity*
- application: Parseval's identity for any ON basis

$$\|v\|_2^2 = v^* v = v^* I v = \sum_{i=1}^{n} v^* q_i q_i^* v = \sum_{i=1}^{n} |\langle q_i, v\rangle|^2$$

spectra of big random matrices

- *claim* (*circular law*). if $A \in \mathbb{R}^{n\times n}$ has entries which are normally-distributed random variables with mean zero and variance n^{-1}, so $a_{ij} \sim N(0, n^{-1})$, then as $n \to \infty$ the spectrum of A fills the unit disc

```
>> A = randn(n,n)/sqrt(n);
>> lam = eig(A);
>> plot(real(lam),imag(lam),'o'),  grid on,  axis([-2 2 -2 2])
```



- but these matrices are not normal

spectra of big random *normal* matrices

- but `randn(n,n)` is not normal (i.e. normal with probablility zero)
- construct a random *normal* matrix with the same spectrum:

```
function [A,B] = gennormal(n);
% GENNORMAL  Generate a random n x n complex matrix A which is normal
% (but not hermitian).  The entries have normal distributions.  The
% eigenvalues will roughly cover the unit disc when n is large.  Also
% returns B, a nonnormal matrix with the same eigenvalues as A.
% Example:
%   >> [A,B] = gennormal(100);
%   >> lam = eig(A);
%   >> plot(real(lam),imag(lam),'o'),  grid on    % same picture for B
%   >> norm(A'*A - A*A')     % very small
%   >> norm(B'*B - B*B')     % not small
% See also GENHERM, PROJMEASURE.

B = randn(n,n)/sqrt(n);  % https://en.wikipedia.org/wiki/Circular_law
                         %     says eigenvalues of B are asymptotically
                         %     uniformly distributed on unit disc
[X,D] = eig(B);          % D is diagonal and holds eigenvalues and
                         %     X holds (nonorthogonal) eigenvectors
[Q,R] = qr(X);           % Q holds ON basis for C^n, built from applying
                         %     orthogonalization to columns of X
A = Q*D*Q';              % construct A to be normal but to have same
                         %     eigenvalues as B
```

spectral subsets correspond to orthogonal projectors

- I also wrote a code `projmeasure.m` which shows $\sigma(A)$ as a subset of $\mathbb{C}$ and lets you select the eigenvalues for which you want eigenvectors
- demo 1:

```
>> A = gennormal(100);
>> P = projmeasure(A);    % <-- user input with mouse
                          %     selects a projector
>> k = rank(P)            % = number of selected eigenvalues
```

- demo 2:

```
>> A = expm(i*eye(6) + gennormal(6));
>> [P,Qh] = projmeasure(A);
>> Qh                     % view selected eigenvectors
```

spectral subsets correspond to orthogonal projectors, cont.

- demo 3:

```
>> U = expm(i*genherm(10));   % random unitary matrix
>> [P,Qh] = projmeasure(U);
>> Qh                         % view selected eigenvectors
```

- John von Neumann ($\sim$1930) imagined this before he invented computers
  - it is a *projector-valued measure*
  - built to handle quantum mechanical operators rigorously
  - he constructed a spectral theorem for normal operators on Hilbert spaces, which uses a projector-valued measure E on $\mathbb{C}$, namely

$$A = \int_{\sigma(A)} \lambda \, dE_\lambda$$

  - the functional calculus:

$$f(A) = \int_{\sigma(A)} f(\lambda) \, dE_\lambda,$$

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

why singular values?

- eigenvalues can be useful!
- but they are only defined for square matrices
  - in ∞-dimensions: "spectrum is useful, but only for $B(X)$, not $B(X, Y)$"
- …and sometimes not so useful anyway
  - only "safe" to use eigenvalues if eigenvectors are orthogonal (A normal)
  - diagonalization $A = S\Lambda S^{-1}$ may tell us little about A when $\kappa(S) \gg 1$
  - square matrices can be defective anyway
- however, *any* $A \in \mathbb{C}^{m\times n}$ has *singular values*
  - what do the eigenvalues say?

    Behavior of powers A^k or functions $f(A)$ like e^{At}.
  - what do the singular values say?

    Invertibility of A: rank, nullity

    Geometric action of A: $\|A\|_2$, $\|A^{-1}\|_2$, condition number, ϵ-pseudospectrum
  - so, what information do you want?

visualizing a matrix

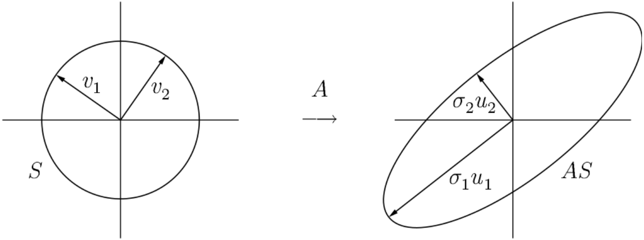


Figure 4.1. *SVD of a* 2×2 *matrix.*

figure from Trefethen & Bau, Numerical Linear Algebra, SIAM Press 1997

- $A \in \mathbb{R}^{m \times n}$ sends the unit sphere in $\mathbb{R}^n$ to a possibly-degenerate hyperellipsoid in $\mathbb{R}^m$
  - the fundamental way to visualize a linear operator
  - also true for $A \in \mathbb{C}^{m \times n}$ . . . but less visualizable
- the *singular values* of A define the geometry of the output hyperellipsoid

singular value decomposition

Theorem

if $A \in \mathbb{C}^{m\times n}$ then there exist $U \in \mathbb{C}^{m\times m}$ unitary, $V \in \mathbb{C}^{n\times n}$ unitary, and $\Sigma \in \mathbb{R}^{m\times n}$ diagonal, with nonnegative entries, so that

$$A = U\Sigma V^*$$

- *singular value decomposition (SVD)* of A
- diagonal entries σ_i of Σ are the *singular values* of A
    - note Σ is same shape as A, while U, V are always square
    - normalization $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min\{m,n\}}$ makes values unique
    - if $A = 0$ we take $\Sigma = A = 0$ and choose U, V as any unitaries
    - if $A \neq 0$ then $\sigma_1 > 0$
- action of $A = U\Sigma V^*$ on a vector:
    - multiplication by V^* finds coefficients of the vector in the columns of V
    - multiplication by Σ stretches the vector along standard axes
    - multiplication by U rotates the vector to the output hyperellipsoid

singular value decomposition: examples

- *example 1.* if $A = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$ then

$$A = \begin{bmatrix} -0.92388 & -0.38268 \\ -0.38268 & 0.92388 \end{bmatrix} \begin{bmatrix} 5.3983 & \\ & 0.92621 \end{bmatrix} \begin{bmatrix} -0.75545 & -0.6552 \\ -0.6552 & 0.75545 \end{bmatrix}^*$$

  - $\|A\|_2 = 5.3983$, $\|A^{-1}\|_2 = 1/0.92621$
  - compare: $\sigma(A) = \{5, 1\}$

- *example 2.* if $B = \begin{bmatrix} 6 & 5 \\ 4 & 3 \\ 1 & 2 \end{bmatrix}$ then

$$B = \begin{bmatrix} -0.82264 & -0.05242 & -0.56614 \\ -0.52578 & -0.30878 & 0.79259 \\ -0.21636 & 0.94969 & 0.22646 \end{bmatrix} \begin{bmatrix} 9.49393 & \\ & 0.93025 \\ & \end{bmatrix} \begin{bmatrix} -0.76421 & -0.64497 \\ -0.64497 & 0.76421 \end{bmatrix}^*$$

  - $\|B\|_2 = 9.49393$
  - B is not invertible
  - $\sigma(B)$ is not defined

singular value decomposition: proof

proof. Induct on n, the column size of A. If $n = 1$ then $A = [a]$ where $a \in \mathbb{C}^m$. Then

$$U = \left[\frac{a}{\|a\|_2}\right], \quad \Sigma = [\|a\|_2], \quad V = [1]$$

is an SVD for A.

For $n > 1$ let $v_1 \in \mathbb{C}^n$ be a unit vector which maximizes the continuous function

$$f(x) = \|Ax\|_2$$

over the compact set $S^n = \{x \in \mathbb{C}^n : \|x\|_2 = 1\}$. (*We just used finite-dimensionality!*) Then Av_1 is a vector in $\mathbb{C}^m$ with length $\sigma_1 = \|Av_1\|_2 = \|A\|_2$. If $\sigma_1 = 0$ we are done because A is the zero matrix. (*Why?*) Otherwise $\sigma_1 > 0$ so let $u_1 = Av_1/\sigma_1$. Now we have $Av_1 = \sigma_1 u_1$.

Extend v_1 and u_1 to orthonormal bases of $\mathbb{C}^n, \mathbb{C}^m$, respectively, giving unitary matrices

$$\tilde{V} = \left[\begin{array}{c|c|c|c} v_1 & \tilde{v}_2 & \dots & \tilde{v}_n \end{array}\right], \qquad \tilde{U} = \left[\begin{array}{c|c|c|c} u_1 & \tilde{u}_2 & \dots & \tilde{u}_m \end{array}\right].$$

Now apply A to $\tilde{V}$,

$$A\tilde{V} = \left[\begin{array}{c|c|c|c} \sigma_1 u_1 & w_2 & \dots & w_n \end{array}\right].$$

Next apply $\tilde{U}^*$, and note that $\tilde{U}^* u_1 = e_1$:

$$\tilde{U}^* A \tilde{V} = \left[\begin{array}{c|c} \sigma_1 & z^* \\ \hline 0 & M \end{array}\right]$$

singular value decomposition: proof cont.

cont. We have

$$\tilde{U}^* A \tilde{V} = \left[\begin{array}{c|c} \sigma_1 & z^* \\ \hline 0 & M \end{array}\right]$$

for $z \in \mathbb{C}^{n-1}$ and $M \in \mathbb{C}^{(m-1)\times(n-1)}$. Because $\tilde{U}$, $\tilde{V}$ are unitary, the matrix norm is unchanged: $\|\tilde{U}^* A \tilde{V}\|_2 = \|A\|_2$.

In fact $z = 0$, as follows. Let $w \in \mathbb{C}^m$ be the vector $w = \begin{bmatrix} \sigma_1 \\ z \end{bmatrix}$. It is nonzero because $\|w\|_2 = (\sigma_1^2 + \|z\|_2^2)^{1/2} \ge \sigma_1 > 0$. But

$$\left\| \left[\begin{array}{c|c} \sigma_1 & z^* \\ \hline 0 & M \end{array}\right] \begin{bmatrix} \sigma_1 \\ z \end{bmatrix} \right\|_2 = \left\| \begin{bmatrix} \sigma_1^2 + z^* z \\ Mz \end{bmatrix} \right\|_2 \ge \sigma_1^2 + \|z\|_2^2 = (\sigma_1^2 + \|z\|_2^2)^{1/2} \|w\|_2.$$

That is, $\|\tilde{U}^* A \tilde{V} w\|_2 \ge (\sigma_1^2 + \|z\|_2^2)^{1/2} \|w\|_2$, so if $z \ne 0$ then $\|A\|_2 = \|\tilde{U}^* A \tilde{V}\|_2 > \sigma_1$, contradicting the definition of σ_1.

Thus

$$\tilde{U}^* A \tilde{V} = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & M \end{array}\right]$$

By the induction hypothesis there exist $\hat{U}, \hat{\Sigma}, \hat{V}$ so that $M = \hat{U} \hat{\Sigma} \hat{V}^*$. Since products of unitaries are unitary, we have an SVD of A:

$$A = \left(\tilde{U} \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \hat{U} \end{array}\right] \right) \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \hat{\Sigma} \end{array}\right] \left(\tilde{V} \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \hat{V} \end{array}\right] \right)^* = U \Sigma V^* \quad \square$$

singular value decomposition: facts

- $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$
- α is a singular value of A if and only if α^2 is an eigenvalue of $A^* A$
- the singular values of A are the same as those of A^*
- for any $A \in \mathbb{C}^{m\times n}$,
    - $\operatorname{rank}(A) = k$ where $\sigma_k > 0$ and $\sigma_{k+1} = 0$
    - $\operatorname{nullity}(A) = \dim(\ker(A)) = q$ where q is the number of zero singular values
- if $A \in \mathbb{C}^{n\times n}$ is square then
    - $|\det(A)| = \prod_{j=1}^n \sigma_j$
    - if A is invertible then $\|A^{-1}\|_2 = 1/\sigma_n$
    - $\kappa_2(A) = \sigma_1/\sigma_n \in [1,\infty]$ is the eccentricity of the output hyperellipsoid
    - $\sigma_n \le \min_{\lambda\in\sigma(A)} |\lambda| \le \max_{\lambda\in\sigma(A)} |\lambda| \le \sigma_1$
- if A is square and normal then $\sigma_j = |\lambda_j|$ (with ordering of $\sigma(A)$)

Outline

1. introduction
2. functional calculus
3. resolvents
4. orthogonal projectors
5. singular value decomposition
6. conclusion

please read the textbook backwards

- go to the end of Chapter 15 “C^* algebras” and read backwards:
  - von Neumann’s spectral theorem for bounded operators on Hilbert spaces
  - functional calculus for normal elements
  - singular value decomposition for compact operators between Hilbert spaces
  - spectral theorem for compact normal operators on a Hilbert space
  - definition of *normal*, *unitary*, and *self-adjoint* (hermitian) elements
  - definition of a C^* algebra
- on the other hand, go to the beginning of Chapter 14 “Spectral theory” and read forward
- I hope that by the end of the semester it will make sense!