18.700 - LINEAR ALGEBRA, DAY 20
SINGULAR VALUE DECOMPOSITION
GENERALIZED EIGENVECTORS

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

I.1. Goals for lesson.

(1) Students will learn Singular Value Decomposition.
(2) Students will learn how to compute an SVD of a matrix.
(3) Students will learn the definition of generalized eigenvector.

I.2. Methods of assessment.

(1) Student responses to questions posed during lecture
(2) Student responses to worksheet

I.3. Materials to bring.

(1) Laptop + adapter (2) Worksheets (3) Chalk

(0:00)

II. LESSON PLAN

II.1. Last time.

- Defined isometries and unitary operators.
- Proved QR Decomposition exists.
- Defined singular values of a linear map.

II.2. 7E: Singular Value Decomposition (SVD), cont.

Definition 1. Let $T \in \mathcal{L}(V, W)$. The *singular values* of T are the nonnegative square roots of the eigenvalues of T^*T, listed in decreasing order, and with multiplicity.

Theorem 2 (Singular Value Decomposition (SVD))**.** *Suppose $T \in \mathcal{L}(V, W)$. Let $s_1, \ldots, s_n$ be the singular values of T, and let $s_1, \ldots, s_m$ be the positive ones. Then there exist orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that*

$$T(v) = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for all $v \in V$.

Proof. Since T^*T is positive, then by the Spectral Theorem there is an orthonormal basis $e_1, \ldots, e_n$ of V such that

$$T^*T(e_k) = s_k^2 e_k$$

for each $k = 1, \ldots, n$. Now, for each $k = 1, \ldots, m$, define

$$f_k := \frac{T(e_k)}{s_k}.$$

We show that $f_1, \ldots, f_m$ is orthonormal:

$$\langle f_j, f_k\rangle = \left\langle \frac{1}{s_j}T(e_j), \frac{1}{s_k}T(e_k)\right\rangle = \frac{1}{s_j s_k}\langle T(e_j), T(e_k)\rangle = \frac{1}{s_j s_k}\langle e_j, T^*T(e_k)\rangle = \frac{1}{s_j s_k}\langle e_j, s_k^2 e_k\rangle$$
$$= \frac{s_k}{s_j}\langle e_j, e_k\rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k \end{cases}$$

for all $j, k \in \{1, \ldots, m\}$. (Note that $s_k \in \mathbb{R}$ for all k.)

Note that for $k > m$ we have

$$T^*T(e_k) = s_k^2 e_k = 0 \implies e_k \in \ker(T^*T) = \ker(T) \implies T(e_k) = 0$$

by a previous result. Given $v \in V$, since $e_1, \ldots, e_n$ is orthonormal, then

$$T(v) = T(\langle v, e_1\rangle e_1 + \cdots \langle v, e_n\rangle e_n = \langle v, e_1\rangle T(e_1) + \cdots \langle v, e_m T(e_m)\rangle$$
$$= \langle v, e_1\rangle s_1 f_1 + \cdots + \langle v, e_m\rangle s_m f_m.$$

□

Proposition 3 (SVD of adjoint)**.** *Suppose $T \in \mathcal{L}(V, W)$ and $s_1, \ldots, s_m$, $e_1, \ldots, e_m$, and $f_1, \ldots, f_m$ are as before, so*

$$T(v) = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for all $v \in V$. Then

$$T^*(w) = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m.$$

Definition 4. Let $A \in M_{m\times n}(\mathbb{F})$. A is *(generalized) diagonl* if $A_{ij} = 0$ for all $i, j \in \{1, \ldots, \min(m, n)\}$ with $i \neq j$.

Theorem 5 (SVD, matrix version). *Let $A \in M_{m\times n}(\mathbb{F})$ have rank r. Then there exists*

- *a generalized diagonal matrix $\Sigma \in M_{\mathrm{m}\times n}(\mathbb{F})$ whose diagonal entries are the positive singular values of A;*
- *a unitary matrix $U \in M_{m\times m}(\mathbb{F})$; and*
- *a unitary matrix $V \in M_{n\times n}(\mathbb{F})$*

such that $A = U\Sigma V^*$.

Proof sketch. Let $v_1, \ldots, v_n$ be an orthonormal basis of $\mathbb{F}^n$ consisting of eigenvectors of A^*A. Let V be the matrix whose columns are $v_1, \ldots, v_n$.

Suppose A has r nonzero singular values. Then $Av_1, \ldots, Av_r$ is an orthogonal basis for $\mathrm{Col}(A)$:

$$\langle Av_i, Av_j\rangle = \langle v_i, A^*Av_j\rangle = \langle v_i, \lambda_j v_j\rangle = \overline{\lambda_j}\langle v_i, v_j\rangle = 0\,.$$

Normalize this list to obtain an orthonormal basis $u_1, \ldots, u_r$ of $\mathrm{Col}(A)$, where

$$u_i := \frac{1}{s_i} Av_i$$

for $i = 1, \ldots, r$. Extend this to a basis $u_1, \ldots, u_m$ of $\mathbb{F}^m$. Let U be the matrix with columns $u_1, \ldots, u_m$. We claim that

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

□

II.3. **Worksheet.**

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \implies A^*A = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}.$$

Then $\mathrm{minpoly}(A^*A) = x^2 - 34x + 225 = (x-9)(x-25)$.

$$V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

$$Av_1 = \begin{pmatrix} 5/\sqrt{2} \\ 5/\sqrt{2} \\ 0 \end{pmatrix} \qquad Av_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 2\sqrt{2} \end{pmatrix}$$

$$u_1 = \frac{1}{5}Av_1$$
$$u_2 = \frac{1}{3}Av_2$$
$$u_3 = \begin{pmatrix} -2/3 \\ 2/3 \\ 1/3 \end{pmatrix}$$

https://sagecell.sagemath.org/?z=eJxtj0FugzAQRfdI3MFL2x0g2M2uqQQ3oBVsLLeyEpqgNqEFy-X4H
vZ_M_MnIztyNrZrBloUoJQEAUKDEiAhElqzMMgQyWLbmUv_2fY1xa_-1H7TDEX-3-PZ1c3RDYOn-GF4RKYolA6DM
5QO5IHh9qS5mnXYqoqz2k0qT_6iwVDMgsxzhOrEPRDfWbkLt0oQXq6m9f5VJwQl8_q5UNKxTloiBNtnzsOSaSoxI
E6DdmC1Avf33Fvzf6dTrM2sNFsauaBuTP3T05v1mHsBwcyiHY=&lang=sage&interacts=eJyLjgUAARUAuQ=
=

Applications of SVD:

- Low-rank approximations of linear maps. Let
$$T(v) = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$
be a singular value decomposition for T. Define $T_k \in \mathcal{L}(V, W)$ as the truncation
$$T_k(v) = s_1\langle v, e_1\rangle f_1 + \cdots + s_k\langle v, e_k\rangle f_k\,.$$
One can show that T_k is the "best" rank k approximation of T. This idea is used in image compression.
- Principal component analysis (PCA). Suppose that we have some multivariate data from n observations, stored as column vectors $X_1, \ldots, X_n \in \mathbb{R}^m$. Let X be the $m \times n$ matrix with columns $X_1, \ldots, X_n$. We want to find an orthonormal basis $v_1, \ldots, v_n$ such that most of the variation of the data occurs in the directions of $v_1, v_2, \cdots v_n$. This is usual done by computing a SVD $X = U\Sigma V^*$: the vectors $v_1, \ldots, v_n$ that are the columns of V (which are also the eigenvectors of X^*X) are the desired basis.

See section 7F for more applications of SVD.

II.4. **8A: Generalized Eigenvectors.** Throughout this section, let V be a nonzero finite-dimensional $\mathbb{F}$-vector space.

II.4.1. *Preliminaries.*

Lemma 6. *Let $T \in \mathcal{L}(V)$. Then we have an ascending chain*
$$\{0\} = \ker(T^0) \subseteq \ker(T) \subseteq \ker(T^2) \subseteq \cdots \subseteq \ker(T^k) \subseteq \ker(T^{k+1}) \subseteq \cdots\,. \tag{7}$$

Proof. [Skip, if necessary.] Suppose $k \in \mathbb{Z}_{\geq 0}$. Given $v \in \ker(T^k)$, then $T^k(v) = 0$. Then
$$T^{k+1}(v) = T(T^k(v)) = T(0) = 0$$
so $v \in \ker(T^{k+1})$. ☐

Proposition 8. *Suppose $T \in \mathcal{L}(V)$. If*
$$\ker(T^m) = \ker(T^{m+1})$$
for some $m \in \mathbb{Z}_{\geq 0}$, then
$$\ker(T^m) = \ker(T^{m+1}) = \ker(T^{m+2}) = \ker(T^{m+3}) = \cdots\,.$$

Proof. [Skip, if necessary.] We want to show that $\ker(T^{m+k}) = \ker(T^{m+k+1})$ for all $k \in \mathbb{Z}_{\geq 0}$. Suppose $k \in \mathbb{Z}_{\geq 0}$.

($\subseteq$): Follows from previous lemma.

($\supseteq$): Given $v \in \ker(T^{m+k+1})$, then
$$T^{m+1}(T^k(v)) = T^{m+k+1}(v) = 0\,,$$

so $T^k(v) \in \ker(T^{m+1}) = \ker(T^m)$. Then

$$T^{m+k}(v) = T^m(T^k(v)) = 0$$

so $v \in \ker(T^{m+k})$. ∎

Proposition 9. *Suppose $T \in \mathcal{L}(V)$. The chain in (7) stabilizes:*

$$\ker(T^n) = \ker(T^{n+1}) = \ker(T^{n+2}) = \cdots,$$

where $n := \dim(V)$.

Proof. For contradiction, suppose $\ker(T^n) \neq \ker(T^{n+1})$. This means that no two terms in the chain are equal at or before the $(n+1)^{\text{st}}$ step by the previous result:

$$\{0\} = \ker(T^0) \subsetneq \ker(T) \subsetneq \cdots \subsetneq \ker(T^n) \subsetneq \ker(T^{n+1})$$

with strict containments. At each strict inclusion, the dimension must increase by at least 1, so $\dim(\ker(T^k)) \geq k$ for each $k = 1, \ldots, n+1$. But then

$$\dim(\ker(T^{n+1})) \geq n+1 > n = \dim(V),$$

contradiction. ∎

It's not true that $V = \ker(T) \oplus \operatorname{img}(T)$ for every $T \in \mathcal{L}(V)$. (Consider $V = \mathbb{F}^3$ and $T := L_A$ where

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

)

Proposition 10. *Suppose $T \in \mathcal{L}(V)$. Letting $n := \dim(V)$, then*

$$V = \ker(T^n) \oplus \operatorname{img}(T^n).$$

Proof. [Skip, if necessary.] Given $v \in \ker(T^n) \cap \operatorname{img}(T^n)$, then $T^n(v) = 0$ and $v = T^n(u)$ for some $u \in V$. Then

$$T^{2n}(u) = T^n(T^n(u)) = T^n(v) = 0,$$

so $u \in \ker(T^{2n}) = \ker(T^n)$. Then

$$0 = T^n(u) = v.$$

Thus $\ker(T^n) \cap \operatorname{img}(T^n) = \{0\}$.

By Rank-Nullity, we have

$$\dim(V) = \dim(\ker(T^n)) + \dim(\operatorname{img}(T^n)) = \dim(\ker(T^n) \oplus \operatorname{img}(T^n)),$$

hence $V = \ker(T^n) \oplus \operatorname{img}(T^n)$. ∎

II.4.2. *Generalized eigenvectors.* Note that v is an eigenvector of $T \in \mathcal{L}(V)$ iff $\operatorname{span}(v)$ is a (1-dimensional) T-invariant subspace. So T is diagonalizable iff

$$V = V_1 \oplus \cdots \oplus V_n \tag{11}$$

where $V_i = \operatorname{span}(v_i)$ is a 1-dimensional T-invariant subspace for each $i = 1, \ldots, n$. But we know that not every linear operator T is diagonalizable: e.g., $V = \mathbb{F}^2$ and $T = L_A$ for

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

But what if we allow for T-invariant subspaces of larger dimension in (11)? This leads to the following notion.

Definition 12. Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is a *generalized eigenvector* of T associated to λ if $v \neq 0$ and

$$(T - \lambda I)^k(v) = 0$$

for some $k \in \mathbb{Z}_{>0}$.

Theorem 13. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.*

Proof. Let $n := \dim(V)$. By strong induction on n.

Base case: $n = 1$. Then every nonzero vector is an eigenvector of T.

Inductive step: Suppose $n \geq 2$ and the result holds for all $k < n$. Since $\mathbb{F} = \mathbb{C}$, then there exists an eigenvalue λ of T. Recall then that

$$V = \ker(T - \lambda I)^n \oplus \operatorname{img}(T - \lambda I)^n.$$

Case 1: $\ker(T - \lambda I)^n = V$. Then every nonzero vector in V is a generalized eigenvector of T, so the result holds.

Case 2: $\ker(T - \lambda I)^n \neq V$. Then $\operatorname{img}(T - \lambda I)^n \neq \{0\}$. Since λ is an eigenvalue of T, then $\dim(\ker(T - \lambda I)^n) \geq 1$. Thus

$$0 < \dim(\operatorname{img}(T - \lambda I)^n) < n.$$

Let $U := \operatorname{img}(T - \lambda I)^n$. Then U is a nonzero T-invariant subspace of dimension $< n$. By the inductive hypothesis applied to the restriction $T|_U$, there is a basis of U consisting of generalized eigenvectors of $T|_U$. Adjoin this basis to a basis of $\ker(T - \lambda I)^n$. Since

$$V = \ker(T - \lambda I)^n \oplus \operatorname{img}(T - \lambda I)^n,$$

then the result is a basis of V consisting of generalized eigenvectors of T. □

Lemma 14. *Suppose $T \in \mathcal{L}(V)$. Then each generalized eigenvector of T corresponds to a unique eigenvalue of T.*

Proof. Exercise. □

Proposition 15. *Suppose $T \in \mathcal{L}(V)$. Then every list of generalized eigenvectors of T corresponding to distinct eigenvalues is linearly independent.*

Proof. Exercise. Similar to the proof for eigenvectors. □

II.4.3. *Nilpotent operators.*

Definition 16. An operator $T \in \mathcal{L}(V)$ is *nilpotent* if $T^m = 0$ for some $m \in \mathbb{Z}_{\geq 0}$.

Example 17. Let $V = \mathbb{F}^2$. The operator $T := L_A$ with

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

is nilpotent.

Proposition 18. *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then* $\operatorname{minpoly}(T) = z^m$ *for some* $m \leq \dim(V)$.

Proof. Since T is nilpotent, then $T^k = 0$ for some $k \in \mathbb{Z}_{\geq 0}$. Then $\operatorname{minpoly}(T)$ divides z^k so

$$z^k = \operatorname{minpoly}(T) f(z)$$

for some $f \in \mathcal{P}(\mathbb{F})$. By unique factorization, then $\operatorname{minpoly}(T) = z^m$ for some $m \in \mathbb{Z}_{\geq 0}$. Moreover, we know that $\deg(\operatorname{minpoly}(T)) \leq \dim(V)$, so $m \leq \dim(V)$. □

Proposition 19. *Let $n := \dim(V)$ and suppose $T \in \mathcal{L}(V)$ is nilpotent. Then $T^n = 0$.*

Proof. By the above, $\operatorname{minpoly}(T) = z^m$ for some $m \leq n$. Then

$$T^n = T^{n-m} T^m = T^{n-m} \circ 0 = 0 \,.$$

□

Proposition 20. *Suppose $T \in \mathcal{L}(V)$.*

(a) If T is nilpotent, then 0 is an eigenvalue of T and T has no other eigenvalues.
(b) If $\mathbb{F} = \mathbb{C}$ and 0 is the only eigenvalue of T, then T is nilpotent.

Proof. (a) By the previous proposition, $\operatorname{minpoly}(T) = T^m$ for some m. The eigenvalues of T are exactly the roots of $\operatorname{minpoly}(T)$.

(b) Since $\mathbb{F} = \mathbb{C}$, then $\operatorname{minpoly}(T)$ splits into degree 1 factors. Then $\operatorname{minpoly}(T) = z^m$ for some $m \in \mathbb{Z}_{\geq 0}$, so $T^m = 0$.

□

Proposition 21. *Suppose $T \in \mathcal{L}(V)$. TFAE.*

(a) T is nilpotent.
(b) $\operatorname{minpoly}(T) = z^m$ *for some* $m \in \mathbb{Z}_{\geq 0}$.
(c) There is a basis $\mathcal{B}$ of V such that

$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & & * \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{pmatrix}$$

where all entries on and below the diagonal are 0.

Proof. (a) $\Longrightarrow$ (b): Already done.

(b) $\Longrightarrow$ (c): Since $\operatorname{minpoly}(T) = z^m$ splits into degree 1 factors, then there is a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is upper triangular. Its diagonal entries are exactly the eigenvalues of T, namely 0, so we obtain a matrix of the desired form.

(c) $\Longrightarrow$ (a): A direct calculation shows that $([T]_{\mathcal{B}})^n = 0$. □