Discussion 6A

The following note is useful for this discussion: Note 16

1. Computing the SVD: A "Tall" Matrix Example

Define the matrix

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}. \tag{1}$$

Here, we expect $U \in \mathbb{R}^{3\times 3}$, $\Sigma \in \mathbb{R}^{3\times 2}$, and $V \in \mathbb{R}^{2\times 2}$ (recall that U and V must be square since they are orthonormal matrices).

In this problem, we will walk through the SVD algorithm, prove some important theorems about the SVD matrices and column/null spaces, and consider an alternate way to approach the SVD.

(a) In this part, we will walk through Algorithm 7 in Note 16. This algorithm applies for a general matrix $A \in \mathbb{R}^{m\times n}$.

i. **Find $r := \text{rank}(A)$. Compute $A^\top A$ and diagonalize it using the spectral theorem (i.e. find V and Λ).**

ii. **Unpack $V := \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ and unpack $\Lambda := \begin{bmatrix} \Lambda_r & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix}$.**

iii. **Find $\Sigma_r := \Lambda_r^{1/2}$ and then find $\Sigma := \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$.**

iv. **Find $U_r := AV_r\Sigma_r^{-1}$, where $U_r \in \mathbb{R}^{3\times 1}$ and then extend the basis defined by columns of U_r to find $U \in \mathbb{R}^{3\times 3}$.**
(HINT: How can we extend a basis, and why is that needed here?)

v. **Use the previous parts to write the full SVD of A.**

vi. **Use the Jupyter notebook to run the code cell that calls `numpy.linalg.svd` on A. What is the result? Does it match our result above?**

Solution: Part **1.**(a)i
We can immediately see that the first column is a multiple of the second, so $r = \text{rank}(A) = 1$. Next, we compute

$$A^\top A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix} \tag{3}$$

The eigenvalues of $A^\top A$ are the roots of $(\lambda - 9)^2 - 81 = 0$, and therefore, $\lambda_1 = 18$ and $\lambda_2 = 0$.

The corresponding orthonormal eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{4}$$

Hence,

$$V = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{5}$$

$$\Lambda = \begin{bmatrix} 18 & 0 \\ 0 & 0 \end{bmatrix} \tag{6}$$

Note that the diagonal entries of Λ are sorted in decreasing order (i.e. $\Lambda_{11} \geq \Lambda_{22}$)

Part **1.**(a)ii

Since $r = 1$, we know that $\Lambda_r \in \mathbb{R}^{1\times 1}$. To unpack it from our computation of Λ, we can pattern match to see that

$$\Lambda_r = \begin{bmatrix} 18 \end{bmatrix} \tag{7}$$

Next, we must find V_r and V_{n-r}. Here, V_r will contain the eigenvectors that correspond to the nonzero eigenvalues, and V_{n-r} will contain all other eigenvectors. Thus,

$$V_r = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{8}$$

$$V_{n-r} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{9}$$

Part **1.**(a)iii

When raising a diagonal matrix to the power $\frac{1}{2}$, we can take the square root of all the diagonal elements. Hence,

$$\Sigma_r = \Lambda_r^{1/2} = \begin{bmatrix} 3\sqrt{2} \end{bmatrix} \tag{10}$$

Now, recall that $\Sigma \in \mathbb{R}^{3\times 2}$. Hence, we can pattern match to the form given in part **1.**(a)iii to obtain

$$\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{11}$$

Part **1.**(a)iv

Plugging in the given formula to find U_r, we obtain:

$$U_r = AV_r\Sigma_r^{-1} = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix} \tag{12}$$

To find U, we must extend the basis formed by the column space of U_r using Gram-Schmidt. This allows us to find two more orthonormal vectors that we can stack in the columns of U to give us an orthonormal matrix. We can do Gram-Schmidt using Python to get

$$\vec{u}_2 = \begin{bmatrix} \frac{\sqrt{8}}{3} \\ \frac{1}{3\sqrt{2}} \\ -\frac{1}{3\sqrt{2}} \end{bmatrix} \qquad \vec{u}_3 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{13}$$

Hence,

$$U = \begin{bmatrix} U_r & \vec{u}_2 & \vec{u}_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{\sqrt{8}}{3} & 0 \\ \frac{2}{3} & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{2}{3} & -\frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{14}$$

Part **1.**(a)v

Finally, we compose this information, and write that A can be decomposed as:

$$A = \underbrace{3\sqrt{2}\begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{compact SVD}} = \underbrace{\begin{bmatrix} -\frac{1}{3} & \frac{\sqrt{8}}{3} & 0 \\ \frac{2}{3} & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{2}{3} & -\frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{full SVD}}. \tag{15}$$

The full SVD representation of A is given below. Note that the full SVD and compact SVD represent the same matrix; the compact form merely omits the columns/rows of U or V which will hit the zero entries of Σ.

Part **1.**(a)vi

The SVD that Jupyter notebook gives is different because of the non-uniqueness of Gram-Schmidt. We can extend a given set of vectors to an orthonormal basis in an infinite number of ways, so the SVD is not unique. Furthermore, it is important to note that the extended columns of U only ever multiply with the zero-entries of Σ. So, they cannot impact the final result of A. However, it is still critical that all the columns of U are in fact mutually orthogonal and normalized.

(b) We now want to create the SVD of $A^\top$. Rather than repeating all of the steps in the algorithm, feel free to use the Jupyter notebook for this subpart (which defines a `numpy.linalg.svd` command). **What are the relationships between the matrices composing A and the matrices composing $A^\top$?**

Solution: We know that A has an SVD representation of $U\Sigma V^\top$ as we solved for above. One natural approach to solving for the SVD of $A^\top$ is to take the transpose of the SVD terms, and "reassign variables". That is, we can say that $A^\top$ has SVD $\widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$ and pattern match to U, Σ, V from before:

$$A^\top = \left(U\Sigma V^\top\right)^\top = V\Sigma^\top U^\top \tag{16}$$

Now, pattern-matching, we can say that $\widetilde{U} = V$, $\widetilde{\Sigma} = \Sigma^\top$, $\widetilde{V}^\top = U^\top \implies \widetilde{V} = U$. Note how the roles have exchanged, and Σ is transposed.

We can write now write the full SVD of $A^\top$ (feel free to confirm that the multiplication yields the right result):

$$A^\top = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{\sqrt{8}}{3} & \frac{1}{3\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{17}$$

(c) **Show, for a general matrix $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$ and $A = U\Sigma V^\top$, that $\text{Null}(A) = \text{Col}(V_{n-r})$. Then, find a basis for the null space of A in eq. (1).** *(HINT: How do we show two sets are equal? Try and use that approach here. Consider the outer product summation form for the SVD. Also, consider using the rank-nullity theorem that* $\dim \text{Col}(A) + \dim \text{Null}(A) = n$*.)*

Solution: $\text{Null}(A) \subseteq \text{Col}(V_{n-r})$

We can start by writing the SVD of A in outer product form:

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top \tag{18}$$

Let us say that we want a (nonzero) vector $\vec{x}$ such that $A\vec{x} = \vec{0}$. This means that

$$A\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top \vec{x} \tag{19}$$

Here, we require $\vec{v}_i^\top \vec{x} = 0$ for $i = 1$ to r. Hence, $\vec{x}$ must be orthogonal to $\vec{v}_i$ for $i = 1$ to r, since each $\sigma_i \neq 0$ and $\vec{u}_i$ are linearly independent. This means that $\vec{x} \in \text{Col}(V_{n-r})$. Since $\vec{x}$ is an arbitrary null space vector of A by construction, it must be the case that $\text{Null}(A) \subseteq \text{Col}(V_{n-r})$.

Having shown this, we can use rank-nullity theorem to see that $\dim \text{Null}(A) = n - r$ since $\text{rank}(A) = r$. Hence, $\dim \text{Null}(A) = \dim \text{Col}(V_{n-r})$ and $\text{Null}(A) = \text{Col}(V_{n-r})$. If you want to go further, you can show the other direction as well:

$\text{Null}(A) \supseteq \text{Col}(V_{n-r})$ (Optional)

If we had an arbitrary $\widetilde{\vec{v}} \in \text{Col}(V_{n-r})$, it must be orthogonal to each of $\vec{v}_i$ for $i = 1$ to r. Hence,

$$A\widetilde{\vec{v}} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \underbrace{\vec{v}_i^\top \widetilde{\vec{v}}}_{=0} = \vec{0} \tag{20}$$

Thus, $\widetilde{\vec{v}} \in \text{Null}(A)$ and $\text{Null}(A) \supseteq \text{Col}(V_{n-r})$.

This completes the proof that $\text{Null}(A) = \text{Col}(V_{n-r})$.

Applying the result above to the A in eq. (1):

$$\text{Null}(A) = \text{Span}\left(\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}\right) \tag{21}$$

(d) **Show, for a general matrix $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$ and $A = U\Sigma V^\top$, that $\text{Col}(A) = \text{Col}(U_r)$. Then, find a basis for the range (or column space) of A.**

Solution: $\mathrm{Col}(A) \subseteq \mathrm{Col}(U_r)$

We can again start by writing the SVD of A in outer product form:

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top \tag{22}$$

We know that $\mathrm{Col}(A) := \{\vec{b} : A\vec{x} = \vec{b}\}$. Hence, if we multiply A by $\vec{x}$:

$$\begin{aligned} A\vec{x} &= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top \vec{x} && (23) \\ &= \sum_{i=1}^{r} \sigma_i \vec{u}_i \left(\vec{v}_i^\top \vec{x}\right) && (24) \\ &= \sum_{i=1}^{r} \left(\vec{v}_i^\top \vec{x}\right) \sigma_i \vec{u}_i && (25) \end{aligned}$$

which is a linear combination of $\vec{u}_1, \ldots, \vec{u}_r$. Hence, $A\vec{x} \in \mathrm{Col}(U_r)$ and so $\mathrm{Col}(A) \subseteq \mathrm{Col}(U_r)$.

Having shown this, we can see that $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(U_r) = r$, so it must be the case that $\mathrm{Col}(A) = \mathrm{Col}(U_r)$. If you want to go further, you can show the other direction as well:

$\mathrm{Col}(A) \supseteq \mathrm{Col}(U_r)$ (Optional)

Suppose $\vec{b} \in \mathrm{Col}(U_r)$, so $U_r\vec{x} = \vec{b}$ for some vector $\vec{x}$. We can show that $\vec{b} \in \mathrm{Col}(A)$. First, we can consider the compact SVD form of A:

$$A = U_r \Sigma_r V_r^\top \tag{26}$$

where $\Sigma_r \in \mathbb{R}^{r \times r}$ is diagonal. Now, we can define the vector $\tilde{\vec{x}} = V_r \Sigma_r^{-1} \vec{x}$. Notice that

$$\begin{aligned} A\tilde{\vec{x}} &= U_r \Sigma_r V_r^\top \tilde{\vec{x}} && (27) \\ &= U_r \Sigma_r V_r^\top V_r \Sigma_r^{-1} \vec{x} && (28) \\ &= U_r \vec{x} = \vec{b} && (29) \end{aligned}$$

so $\vec{b} \in \mathrm{Col}(A)$, and thus, $\mathrm{Col}(A) \supseteq \mathrm{Col}(U_r)$.

This completes the proof that $\mathrm{Col}(A) = \mathrm{Col}(U_r)$.

In this specific case for the matrix A in eq. (1):

$$\mathrm{Col}(A) = \mathrm{Span}\left(\left\{\begin{bmatrix} -1/3 \\ 2/3 \\ -2/3 \end{bmatrix}\right\}\right) \tag{30}$$

(e) **(PRACTICE) Show, for a general matrix $A \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}(A) = r$ and $A = U\Sigma V^\top$, that $\mathrm{Null}(A^\top) = \mathrm{Col}(U_{m-r})$ and $\mathrm{Col}(A^\top) = \mathrm{Col}(V_r)$. Then show:**

i. $\dim \mathrm{Col}(A) + \dim \mathrm{Null}(A^\top) = n$**,**

ii. and $\mathrm{Col}(A)$ and $\mathrm{Null}(A^\top)$ are orthogonal.

Solution: Combining the results from the previous parts and using the notation from part **1.b**, we know $\mathrm{Null}(A^\top) = \mathrm{Col}\left(\widetilde{V}_{m-r}\right) = \mathrm{Col}(U_{m-r})$ and $\mathrm{Col}(A^\top) = \mathrm{Col}\left(\widetilde{U}_r\right) = \mathrm{Col}(V_r)$.

We know that $m = \dim \mathrm{Col}(U_r) + \dim \mathrm{Col}(U_{m-r}) = \dim \mathrm{Col}(A) + \dim \mathrm{Null}(A^\top)$. Since $\mathrm{Col}(A) = \mathrm{Col}(U_r)$ and $\mathrm{Null}(A^\top) = \mathrm{Col}(U_{m-r})$, $\mathrm{Col}(A)$ and $\mathrm{Null}(A^\top)$ must be orthogonal since all the columns of U_r and all the columns of U_{m-r} are orthogonal.

(f) Suppose A was a wide matrix. Instead of finding $A^\top A$, we may want to find the SVD by computing $AA^\top$. The original Algorithm 7 from Note 16, in its entirety, is shown below:

Algorithm 1 Constructing the SVD

function FULLSVD($A \in \mathbb{R}^{m\times n}$)
  $r :=$ RANK(A)
  $(V, \Lambda) :=$ DIAGONALIZE$(A^\top A)$ ▷ Sorted so that $\Lambda_{11} \geq \cdots \geq \Lambda_{nn}$
  Unpack $V := \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$
  Unpack $\Lambda := \begin{bmatrix} \Lambda_r & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix}$
  $\Sigma_r := \Lambda_r^{1/2}$
  Pack $\Sigma := \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$
  $U_r := AV_r\Sigma_r^{-1}$
  $U :=$ EXTENDBASIS$(U_r, \mathbb{R}^m)$
  return (U, Σ, V)
end function

Write a modified version of Algorithm 7 where you solve for the SVD of A using $AA^\top$ instead of $A^\top A$. *(HINT: Consider replacing every instance of "A" in $A^\top A$ with "$A^\top$". What happens? How can we use the result from the* ***1.b*** *part?)*

Solution:

Algorithm 2 Constructing the SVD

function FULLSVD($A \in \mathbb{R}^{m\times n}$)
  $r :=$ RANK(A)
  $(U, \Lambda) :=$ DIAGONALIZE$(AA^\top)$ ▷ Sorted so that $\Lambda_{11} \geq \cdots \geq \Lambda_{nn}$
  Unpack $U := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$
  Unpack $\Lambda := \begin{bmatrix} \Lambda_r & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix}$
  $\Sigma_r := \Lambda_r^{1/2}$
  Pack $\Sigma := \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$
  $V_r := A^\top U_r\Sigma_r^{-1}$
  $V :=$ EXTENDBASIS$(V_r, \mathbb{R}^n)$
  return (U, Σ, V)
end function

Contributors:

- Anish Muthali.
- Neelesh Ramachandran.
- Druv Pai.
- John Maidens.