Linear Algebra and Optimization
MIT Fall 2022

Recitation 12

Thursday October 20, 2022

1 Recap

1.1 Orthogonality

A matrix $A \in \mathbb{R}^{m \times n}$ has orthonormal columns iff $A^\top A = I_n$.
If in addition, A is square with $m = n$, we also have $AA^\top = I_n$ and $A^{-1} = A^\top$. In this case, we say that A is an *orthogonal matrix*.

1.2 Singular Values Decomposition (SVD)

Let A be an $n \times m$ matrix. Then, there exist a factorization $A = U\Sigma V^\top$ where

1. Dimensions: U is $n \times n$, Σ is $n \times m$, V is $m \times m$.
2. U and V are orthogonal matrices.
3. Σ is nonnegative and diagonal (as much as possible, since in general it is rectangular); that is all non-diagonal entries must be 0, and the diagonal entries must be non-negative. Usually Σ is written in terms of $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = 0$ where σ_i is the diagonal entry in the i^{th} row.

Another way to express the decomposition is

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$$

where u_i is the i^{th} column of U and v_i is the i^{th} column of V. The σ_i's are called *singular values*, and the u_i's and v_i's are left and right *singular vectors*, respectively.

1.3 SVD Properties

1. The rank of A is equal to r, which is the number of non-zero singular values
2. The vectors $\{u_1, \ldots, u_r\}$ are an orthonormal basis of $C(A)$.
   - Each u_i is in $C(A)$ because $u_i = A(v_i/\sigma_i)$. They are linearly independent, and there are r of them which is equal to the dimension of $C(A)$.
3. The vectors $\{v_{r+1}, v_{r+1}, ..., v_m\}$ are an orthonormal basis of $N(A)$.
   - Each v_j is in $N(A)$ because $Av_j = 0$. They are linearly independent, and there are $m - r$ of them which is equal to the dimension of $N(A)$.

1.4 SVD and Matrix Inverses

Suppose a matrix A has an SVD: $A = U\Sigma V^\top$.

1. If A is invertible, then $A^{-1} = V\Sigma^{-1}U^\top$. This is because

$$\begin{aligned}(V\Sigma^{-1}U^\top)A &= (V\Sigma^{-1}U^\top)(U\Sigma V^\top) = V\Sigma^{-1}(U^\top U)\Sigma V^\top \\ &= V\Sigma^{-1}\Sigma V^\top = V(\Sigma^{-1}\Sigma)V^\top = VV^\top = I\end{aligned}$$

2. If A is not invertible, we define the *pseudoinverse* of A, denoted A^+, to be

$$A^+ = \sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^\top.$$

   Some crucial properties of pseudoinverses include 1) AA^+ is the orthogonal projection matrix onto $C(A)$, and 2) A^+A is the orthogonal projection matrix onto $N(A)^\perp$.

   Oftentimes a pseudoinverse can be a great substitution of the (hypothetical) inverse for some functionalities. Note that we can create pseudoinverses of rectangular inverses, although only square matrices are invertible.

2 Exercises

1. Let $A \in \mathbb{R}^{m\times n}$ with SVD $A = U\Sigma V^\top$ with rank r. Write down an SVD for the following matrices. What are their non-zero singular values in terms of $\sigma_1, ..., \sigma_r$?
   (a) cA when c is a scalar
   (b) $A^\top$
   (c) $AA^\top$
   (d) A^+

2. Let A be a square matrix and invertible. Explain why $A^+ = A^{-1}$. *Hint: we can explain it both algebraically and geometrically. You may want to use the uniqueness properties of the SVD.*

3. Can you write an SVD of the following matrix A by hand?

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

   Compute the pseudoinverse of the following matrix. We haven't learned yet how to compute a pseudoinverse, so you'll have to improvise here! *Hint: Can you use the orthonormal nullspace and*

4. Show that for *every* matrix A, the following properties hold:
   (a) $AA^+A = A$
   (b) $A^+AA^+ = A^+$
   (c) Both AA^+ and A^+A are symmetric matrices.

   It can be shown that these properties *uniquely* define the pseudoinverse (sometimes also called the Moore-Penrose inverse).

3 Solutions

1. (a) $cA = U(c\Sigma)V^\top$ – that is the singular values are multiplied by c.

   (b) $A^\top = \left(U\Sigma V^\top\right)^\top = V\Sigma^\top U$ which is an SVD because the dimensions match, both V and U are orthonormal matrices, and $\Sigma^\top$ is nonnegative and diagonal. The singular values are those values in diagonal entries of $\Sigma^\top$ which are the same as Σ's – which are $\sigma_1, ..., \sigma_r$.

   (c) $AA^\top = (U\Sigma V^\top)(V\Sigma^\top U^\top) = U\Sigma(V^\top V)\Sigma^\top U^\top = U(\Sigma\Sigma^\top)U^\top$ which is an SVD form. Its singular values are those values in diagonal entries of $\Sigma\Sigma^\top$ which are $\sigma_1^2, ..., \sigma_r^2$. Furthurmore, it tells us that $AA^\top$ has r non-zero singular values which means $\text{rank}(AA^\top) = r = \text{rank}(A)$.

   (d) Recall that $A^+ = \sum_{i=1}^r \sigma_i^{-1} v_i u_i^\top$ which can also be expressed as $V\Sigma^+ U^\top$ where Σ^+ is $\Sigma^\top$ with a non-zero diagonal entry replaced with its reciprocal. This means A^+'s singular values are $\sigma_1^+ = 1/\sigma_1, \sigma_2^+ = 1/\sigma_2$,..., $\sigma_r^+ = 1/\sigma_r$ and 0 otherwise.

2. Algebraic explanation: If A is invertible, then $r = n$ which means none of the singular values is 0. Recall the question 1d. If Σ is $n \times n$ and none of the diagonal entries is 0, so is Σ^+. It follows that $\Sigma\Sigma^+ = I_n$ which means $\Sigma^+ = \Sigma^{-1}$. This means (from 1d) that $A^+ = V\Sigma^+ U^\top = V\Sigma^{-1}U^\top = A^{-1}$.

   Geometric explanation: If A is invertible, then $C(A) = \mathbb{R}^n$. This means any vector $x \in \mathbb{R}^n$ has the projection onto $C(A)$ to be x itself – as it is already in $C(A)$. In other words, the orthogonal projection matrix onto $C(A) = I_n$.

   On the other hand, we know that AA^+ is the orthogonal projection matrix onto $C(A)$. This means $AA^+ = I_n$ which implies $A^+ = A^{-1}$.

3. We observe that $\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}0 & 1 & 0\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$. Therefore, $A = \sigma_1 u_1 v_1^\top = U\Sigma V^\top$ when $U = \begin{bmatrix}1 & ?\\0 & ?\end{bmatrix}$, $\Sigma = \begin{bmatrix}-2 & 0 & 0\\0 & 0 & 0\end{bmatrix}$, and $V = \begin{bmatrix}0 & ? & ?\\1 & ? & ?\\0 & ? & ?\end{bmatrix}$. Note that the "?"s are whatever values needed to make U and V orthogonal. Their exact numerical values do not matter since the only non-zero singular value is $\sigma_1 = -2$. We can notice further that the singular value is negative (-2). This is not a major issue; we can resolve it by re-writing $U = \begin{bmatrix}-1 & ?\\0 & ?\end{bmatrix}$, $\Sigma = \begin{bmatrix}2 & 0 & 0\\0 & 0 & 0\end{bmatrix}$, and $V = \begin{bmatrix}0 & ? & ?\\1 & ? & ?\\0 & ? & ?\end{bmatrix}$.

   It follows that $A^+ = \sigma_1^{-1} v_1 u_1^\top = (1/2)\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}-1 & 0\end{bmatrix} = \begin{bmatrix}0 & 0\\-1/2 & 0\\0 & 0\end{bmatrix}$.

4. We know that $A^+ = V\Sigma^+ U^\top$, where Σ^+ is $\Sigma^\top$ with a non-zero diagonal entry replaced with its reciprocal.

   For the first two parts, it suffices to show that the identities hold for Σ and Σ^+ instead of A and A^+, since for example

   $$AA^+A = U\Sigma V^\top V\Sigma^+ U^\top U\Sigma V^\top = U\Sigma\Sigma^+\Sigma V^\top.$$

Suppose Σ is rank r, so it can be written as $\Sigma = \begin{bmatrix} D & 0_{r\times q} \\ 0_{p\times r} & 0_{p\times q} \end{bmatrix}$ where D is an invertible diagonal matrix. Thus $\Sigma^+ = \begin{bmatrix} D^{-1} & 0_{r\times p} \\ 0_{q\times r} & 0_{q\times p} \end{bmatrix}$, $\Sigma^+\Sigma = \begin{bmatrix} I_r & 0_{r\times q} \\ 0_{q\times r} & 0_{q\times q} \end{bmatrix}$ and $\Sigma\Sigma^+ = \begin{bmatrix} I_r & 0_{r\times p} \\ 0_{p\times r} & 0_{p\times p} \end{bmatrix}$. Then it is easy to see that $\Sigma\Sigma^+\Sigma = \Sigma$ and $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$. For the last part, $AA^+ = U\Sigma\Sigma^+U^\top$ and $A^+A = V\Sigma^+\Sigma V^\top$, and the result follows from $\Sigma\Sigma^+$ and $\Sigma^+\Sigma$ both being symmetric matrices.