SVD and the Psuedoinverse (HW #8)

This material is in Chapter 7 of the 5th edition of your textbook. I will test and quiz you on the parts of sections 7.2 and 7.4 that I cover in these notes. I'm leaving out a lot of material! The rest of Chapter 7 is for you to enjoy.

Every m x n matrix A with real entries can be factored as $A = U\Sigma V^T$ where U and V are orthogonal matrices, and so of size m x m and n x n respectively, and Σ is an m x n diagonal matrix. This is the **Singular Value Decomposition** or **SVD**. The **nonzero** entries in Σ are called the **singular values** of A. The factorization looks like this where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$:

$$A = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_r & \vec{u}_{r+1} & \dots & \vec{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & \text{\Large 0} & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & \text{\Large 0} & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_r & \vec{v}_{r+1} & \dots & \vec{v}_n \end{bmatrix}^T$$

Find a basis for the four fundamental spaces of A. It is easy to see what the bases for the null space and left null space are once you have written the SVD in rank-1 form, and then the orthogonality of U and V with the help of the FTLA gives us the basis for the column and row spaces.

Given a matrix A, how do we find the corresponding SVD?

We start by finding the eigenvalues for A^TA or AA^T, whichever is smaller. Let's suppose A^TA is the smaller matrix. It is positive semi-definite [1], so the eigenvalues are non-negative and the Spectral theorem guarantees us an orthonormal basis of eigenvectors.

Since $N(A) = N(A^TA)$ [2] and both matrices have n columns, the rank r of A equals the rank of A^TA. Define the σ_i^2 to be the i^{th} positive eigenvalue with a corresponding **unit** eigenvector $\vec{v}_i$ for $i = 1, 2, ..., r$.

Using Gram-Schmit if necessary, construct an orthonormal basis of $N(A)$ and name it $\vec{v}_{r+1}, ..., \vec{v}_n$. Similarly, construct an orthonormal basis of $N(A^T)$ and call it $\vec{u}_{r+1}, ..., \vec{u}_m$.

Since we are constructing $A = U\Sigma V^T$ then $AV = U\Sigma$ so we must have $A\vec{v}_i = \vec{u}_i\sigma_i$ or

$$\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}.$$

We will talk about why this last formula is true later. For now, let's use it to help construct an SVD.

Find the SVD for $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 2 \\ 1 & 1 & -2 \\ 0 & 0 & 2 \end{bmatrix}$.

[1] $\vec{x}^TA^TA\vec{x} = \|A\vec{x}\|^2 \geq 0$

[2] $x \in N(A^TA) \implies \vec{x}^TA^TA\vec{x} = 0 \implies \|A\vec{x}\|^2 = 0 \implies A\vec{x} = \vec{0} \implies \vec{x} \in N(A)$. Also, $\vec{x} \in N(A) \implies A^TA\vec{x} = \vec{0} \implies \vec{x} \in N(A^TA)$.

On tests and quizzes, I will only ask you to construct the rank-1 form so that you can avoid finding the basis for the null spaces. What is the rank-1 form for the last example?

Let's examine the formula $\vec{u}_i = \dfrac{A\vec{v}_i}{\sigma_i}$. Does it give an orthonormal set of r vectors that span the column space of A? The $\vec{u}_i$ are certainly in $C(A)$ since they equal a linear combination of the columns of A.

Can you prove $\vec{u}_i = \dfrac{A\vec{v}_i}{\sigma_i}$ is a unit vector? Hint: prove $\|\vec{u}_i\|^2 = 1$.

Can you prove that $\vec{u}_i$ and $\vec{u}_j$ are orthogonal if $i \neq j$? Hint: prove the dot product is zero. Remember $\vec{v}_i$ and $\vec{v}_j$ are orthogonal.

Show that $\vec{u}_i = \dfrac{A\vec{v}_i}{\sigma_i}$ is an eigenvector of AA^T with eigenvalue σ_i^2. Hint: $\vec{v}_i$ is an eigenvector for $A^T A$.

So we could have found the $\vec{u}_i$ first using AA^T. If we start this way, what formula will give us $\vec{v}_i$?

Find the rank-1 decomposition of the SVD for $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

Psuedoinverse

If $A = U\Sigma V^T$ is the SVD of A, and A is invertible, then what is A^{-1} in terms of U, Σ, and V? Give an example of Σ^{-1}.

If A is not square, say m x n where m $\neq$ n, then how should we define a psuedo inverse for A?

The **Psuedoinverse** of m x n A with SVD $A = U\Sigma V^T$ is the n x m matrix

$$A^+ = V\Sigma^+U^T$$

where $\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & & & & & \\ & \sigma_2^{-1} & & & \mathbf{0} & & \\ & & \ddots & & & & \\ & & & \sigma_r^{-1} & & & \\ & & & & 0 & & \\ & \mathbf{0} & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$ is an n x m matrix.

Find A^+ if $A = 4\frac{\begin{bmatrix}1\\1\\-1\\1\end{bmatrix}}{2}\begin{bmatrix}0 & 0 & 1\end{bmatrix} + 2\frac{\begin{bmatrix}1\\0\\1\\0\end{bmatrix}}{\sqrt{2}}\frac{\begin{bmatrix}1 & 1 & 0\end{bmatrix}}{\sqrt{2}}$ is the SVD of A in rank-1 form.

Psuedo Theorem: If $A\vec{x} = \vec{b}$, then $\vec{x}^+ = A^+\vec{b}$ is the "best solution" - it is the Fourier coefficient of a projection with smallest magnitude.

When the columns of A are linearly independent, then $\vec{x}^+ = \hat{x}$ that we calculate when finding projections. Otherwise, it gives the best $\hat{x}$ of all the infinite choices.

Find the "best solution" to $A\vec{x} = \vec{b}$ if $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and $\vec{b} = (1, 2, -1, -2)$. Notice the system is unsolvable and there will be many $\hat{x}$'s since the columns of A are dependent.

Here is a sketch of the proof for the Psuedo Theorem, **only for those who are interested**.

Recall that the projection of $\vec{b}$ onto a **unit** vector $\vec{u}$ is

$$\frac{\vec{u}^T\vec{b}}{\vec{u}^T\vec{u}}\vec{u} = \vec{u}\frac{\vec{u}^T\vec{b}}{1} = \left(\vec{u}\vec{u}^T\right)\vec{b}.$$

Since the $\vec{u}_i$ are orthonormal, the projection onto the column space they span is $\sum_{i=1}^{r}\left(\vec{u}_i\vec{u}_i^T\right)\vec{b}$. But if $\vec{x}^+ = A^+\vec{b}$, then

$$A\vec{x}^+ = AA^+\vec{b} = \sum_{i=1}^{r}\left(\vec{u}_i\vec{u}_i^T\right)\vec{b}.$$

That last equality follows from $AA^+ = (U\Sigma V^T)(V\Sigma^+U^T)$, and so $\vec{x}^+$ is indeed an $\hat{x}$ - or Fourier coefficient.

If $\hat{x}$ is any other Fourier coefficient, then $A^TA\hat{x} = A^TA\vec{x}^+$ implies $A^TA(\vec{x}^+ - \hat{x}) = \vec{0}$, so $\hat{x} = \vec{x}^+ + \vec{n}$ where $\vec{n} \in N(A^TA) = N(A)$. Then

$$\|\hat{x}\|^2 = \|\vec{x}^+\|^2 + 2\vec{x}^+\cdot\vec{n} + \|\vec{n}\|^2.$$

However,

$$\vec{x}^+ = A^+\vec{b} = \sum_{i=1}^{r}\frac{1}{\sigma_i}\vec{v}_i\left(\vec{u}_i^T\vec{b}\right)$$

is a linear combination of the $\vec{v}_i$, so it is in the row space of A implying $\vec{x}^+\cdot\vec{n} = 0$. It then must follow that

$$\|\hat{x}\|^2 = \|\vec{x}^+\|^2 + \|\vec{n}\|^2 \implies \|\hat{x}\| \geq \|\vec{x}^+\|$$

as we hoped to prove.

Chapter 7 has a good general discussion of how the SVD is applied. To get an idea of how much it is used, go to https://arxiv.org - a website provided by Cornell University listing research articles - and search for SVD. For instance, my recent search returned 350 articles that included SVD in the abstract from February 2017 to the present (March 28, 2020).