Pseudoinverse, Change of Basis (cont.), Linear Transformation [*] Perron-Frobenius Theorem

1. **The Pseudoinverse**
   Last time we arrived at $A = U\Sigma V^T$, where U and V are the eigenvectors of AA^T and $A^T A$ so that

$$Av_j = \begin{cases} \sigma_j u_j & \text{for } j \leq r \\ 0 & \text{for } j \geq r \end{cases} \tag{1}$$

We go from row space to the column space for $j \leq r$. We do the reverse: go from column space back to row space:

$$A^+ u_j = \begin{cases} \frac{1}{\sigma_j} v_j & \text{for } j \leq r \\ 0 & \text{for } j \geq r \end{cases} \tag{2}$$

Let's write the above equation into matrices:

$$A^+ = V\Sigma^+ U^T \tag{3}$$

The A^+ above is just called the pseudoinverse of A.
Easy Exercise (Strang 7.4-13) If A has rank r, does A^+ also have rank r?
Exercise: Do it yourself (Strang 7.4-10, 11) Obtain the SVD of $A = \begin{bmatrix} 3 & 4 & 0 \end{bmatrix}$. Find its pseudoinverse.

(a) **The difference with the normal inverse**
When A is invertible, the second case for the pseudoinverse does not exist, its pseudoinverse is just the inverse; its inverse is just the pseudoinverse. (You can see it by simply taking the inverse on both sides of $A = U\Sigma V^T$). What the inverse does is that it simply maps a vector from its column space to its row space. The inverse's column space is the row space of the original matrix.
Now, when the matrix is not invertible, left nullspace and nullspace are not **Z** anymore. What the pseudoinverse does in addition to what an inverse does to vectors in the column space is that it kills any vectors from the left nullspace to the zero vector. This is the second case in the equation 2.

(b) AA^+ and A^+A
AA^+ projects onto the column space of A;
A^+A projects onto the row space of A;
Exercise Find AA^+ and A^+A of the A in the previous exercise.
To see why they are projection matrix, you can refer to Strang 7.4-21 for arriving at it analytically. I include it below nevertheless so that you see it.

$$A^+ A = \left(\sum_1^r \frac{v_i u_i^T}{\sigma_i}\right)\left(\sum_1^r \sigma_j u_j v_j^T\right) = \sum_1^r v_i u_i^T u_i v_i^T = \sum_1^r v_i v_i^T \tag{4}$$

the second equation is because $u_i^T u_j = 0$ when $i \neq j$. Now it is clear that it is a (sum of) projection(s). Similarly, $AA^+ = \sum_1^r u_i u_i^T$.

---

[*] One thing that I will not review today is the matrix norm. If you need help, the textbook is pretty clear about the concepts.

(c) **Application to least squares with dependent columns**
Reading: Chapter 7.4 "Application to least squares with dependent columns"
Here let's briefly see why $\hat{\boldsymbol{x}} = A^+\boldsymbol{b}$ is the shortest solution to the problem

$$A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \tag{5}$$

Observe the equation on the right side: $A^T\boldsymbol{b}$, we can decompose $\boldsymbol{b}$ into a vector in the column space of A, $\boldsymbol{b}_{\text{column}}$ and a vector in the left nullspace of A, $\boldsymbol{b}_{\text{left\_null}}$. Let's first consider the case of just $\boldsymbol{b} = \boldsymbol{b}_{\text{left\_null}}$ or $\boldsymbol{b} = \boldsymbol{b}_{\text{column}}$.

  i. When $\boldsymbol{b} = \boldsymbol{b}_{\text{left\_null}}$
     We get $A^T\boldsymbol{b} = \boldsymbol{0}$ on the right side of the equation. We thus want $A\hat{\boldsymbol{x}} = 0$. This is saying that $\hat{\boldsymbol{x}}$ is in the nullspace of A. If we are after the shortest $\hat{\boldsymbol{x}}$, then $\hat{\boldsymbol{x}} = \boldsymbol{0}$.

  ii. When $\boldsymbol{b} = \boldsymbol{b}_{\text{column}}$
     We get $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}_{\text{column}}$, which is saying $A\hat{\boldsymbol{x}} = \boldsymbol{b}_{\text{column}}$, $\hat{\boldsymbol{x}}$ will then have to be in the rowspace of A so that $\boldsymbol{b}_{\text{column}} \neq \boldsymbol{0}$.
     Combining the two cases, you see a parallelism to equation 2. $\hat{\boldsymbol{x}} = A^+\boldsymbol{b}$.

When $\boldsymbol{b}$ is a random vector, decompose it into the 2 cases, and the sum of the two cases will satisfy the equation 5.

Exercise (Strang 7.4 A) If A has full column rank, then it has a left inverse $C = (A^T A)^{-1}A^T$ that gives $CA = I$. Explain why the pseudoinverse is $A^+ = C$ in this case.

2. **Change of basis and Similar Matrices (cont.)**
We knew before (I already introduced this topic in 6.1-6.2) that changing an input basis from the standard basis to an output basis put in the columns of M can be realized by the matrix M^{-1} (from standard to M). On the other hand, M can change the output from using the basis in M to standard.

Exercise What matrix changes input basis from V to that of W's (bases in the columns of V and W)?

Next, we derive the formula for any transformation in a new basis set $\boldsymbol{m}_1, ..., \boldsymbol{m}_n$:

$$B_{\boldsymbol{m}\text{'s to }\boldsymbol{m}\text{'s}} = M^{-1}_{\text{standard to }\boldsymbol{m}\text{'s}}A_{\text{standard}}M_{\boldsymbol{m}\text{'s to standard}} \tag{6}$$

Note that the transformation by A and B's are the **same** transformation—just reads input and spits out output in a different basis set than the standard's.
Exercise Use eigenvectors as your new basis set, B is now what matrix?

3. **The Geometry behind $A = U\Sigma V^T$**
From the "change of basis" point of view, we get

$$A_{\text{standard to standard}} = U_{\boldsymbol{u}\text{'s to standard}}\Sigma_{\boldsymbol{v}\text{'s to }\boldsymbol{u}\text{'s}}V^T_{\text{standard to }\boldsymbol{v}\text{'s}} \tag{7}$$

Alternatively,

$$\Sigma_{\boldsymbol{v}\text{'s to }\boldsymbol{u}\text{'s}} = U^T_{\text{standard to }\boldsymbol{u}\text{'s}}A_{\text{standard to standard}}V_{\boldsymbol{v}\text{'s to standard}} \tag{8}$$

You can think of equation 7 as a rotation to another basis direction, then stretching (output into another direction), then a final rotation back to the standard basis direction. The next decomposition, called the Polar Decomposition separates rotation and stretching.
The following 2 exercises shall help you get a sense of a transformation from $\boldsymbol{v}$ to $\boldsymbol{u}$.

Exercise (Strang 8.2-5) With bases v_1, v_2, v_3 and w_1, w_2, w_3, suppose $T(v_1) = w_2$, $T(v_2) = T(v_2) = w_1 + w_3$, what is the matrix $A_{v\text{'s to }w\text{'s}}$ for the transformation.

Exercise (Strang 8.1-12) Suppose a linear transformation T transforms (1,1) to (2,2) and (2,0) to (0,0). Find $T(v)$ when $v = (3,1)$ and in general $v = (v_1, v_2)$.

4. **Polar Decomposition**
   Polar decomposition: $A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QH = (U\Sigma U^T)(UV^T) = KQ$
   H, K is positive semidefinite in general, and $H^2 = A^T A, K^2 = AA^T$.

5. **Linear Transformation**

   (a) The transformation T acting on each input vector in V and W is linear if $T(cv) = cT(v)$ for all c and $T(v + w) = T(v) + T(w)$. Equivalently, $T(cv + dw) = cT(v) + dT(w)$ for all c and d.

   (b) Some terminologies:
   (1) **Range**: set of all outputs $T(v)$ (2) **Kernel**: set of all inputs $T(v) = 0$ You will soon see how all linear transformation can be produced by matrices: If $T(v) = Av$, range is column space and kernel is nullspace.

   Exercise (Strang 8.1-Example 6) Is the transformation linear: Project every 3-dimensional vector onto the horizontal plane $z = 1$. Can you write a matrix A for describing this transformation?

   Exercise (Strang 8.1-Example 7) Suppose A is an invertible matrix, is the transformation linear?

6. **Use matrix to produce linear transformation**
   For the basis set $v_1, v_2...v_n$, we have $T(v_1), T(v_2), ...T(v_n)$ as columns of A. This A is the $A_{v\text{'s to }v\text{'s}}$
   Exercise Deduce the rotation matrix in the xy plane.

7. **Matrix products matches transformations**
   You can think of equation 6 as $\mathbf{I \times T}$ **(a transformation done by matrix** A**)$\times$I**, as "change of basis" only changes basis while keeping the identity.

8. **Perron-Frobenius Theorem**
   You will see this theorem in various contexts, for example, in your course in probability. The proof will be skipped here. Let's state this theorem (though it is still only part of it):
   Suppose A has all entries greater than 0, A has a positive (real) eigenvalue $\lambda_{\max}$ such that all other eigenvalues of A satisfy $|\lambda| \leq \lambda_{\max}$. Furthermore, $\lambda_{\max}$ has algebraic and geometric multiplicity one, and has an eigenvector x with all elements in $x > 0$.

   More Exericise 1 (Strang 8.2 20,21) One basis for second-degree polynormials is $v_1 = 1$, $v_2 = x$, $v_3 = x^2$. Another basis is $w_1 = \frac{1}{2}(x^2 + x)$, $w_2 = 1 - x^2$, $w_3 = \frac{1}{2}(x^2 - x)$. Find the two change of basis matrices, from the w's to the v's, and from the v's to the w's.

   Do it yourself: Worked Example 8.2A and 8.2-33, **8.2-3&8** (Great exercises for showing your when A^2. My hint: add in subscript of A.), **10&11** (you'd see a closer connection between linear transformation and matrices), **22** (a refresher on matrix notation), **24** (Understand QR decomposition in terms of "change of basis"), **26**, **27&32** (A fun problem).

Next time will be our **final** class! In addition to chapter 8.3, we would do some review by talking about (1) "Elipse" of $\boldsymbol{x}^T A \boldsymbol{x}$ and $A\boldsymbol{x}$ (review Strang 6.5 and 7.4-14); (2) When are eigenvectors are normal? (3)...and maybe something with positive definite matrices by introducing you the central matrix in Spectral Clustering! Stay tuned. (This also means next class might be longer than 45 min.)