Required Problems

1. Given an $m \times n$ matrix A with (full format) singular value decomposition $A = U\Sigma V^T$, with

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

define Σ^+ to be the $n \times m$ matrix

$$\Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & 1/\sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Define the **pseudoinverse** of A, denoted A^+, to be $A^+ = V\Sigma^+U^T$.

(a) Compute $\Sigma^+\Sigma$. How many rows and columns does it have?
(b) Compute $\Sigma\Sigma^+$. How many rows and columns does it have?
(c) How many rows and columns does A^+ have?
(d) Show that $AA^+ = U_rU_r^T$.
(e) Prove that if A has full row rank, $AA^+ = I_m$.
(f) Show that $A^+A = V_rV_r^T$.
(g) Prove that if A has full column rank, $A^+A = I_n$.
(h) Prove that if A is invertible, $A^+ = A^{-1}$.

Solution:

(a) $$\Sigma^+\Sigma = \begin{bmatrix} 1/\sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & 1/\sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & 0 \\ & & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Where The result is an nxn matrix with n rows and n columns

(b) $$\Sigma\Sigma^+ = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & 1/\sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & 0 \\ & & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Where The result is an mxm matrix with m rows and m columns

(c) A^+ is an n x m matrix with n rows and m columns

(d) Begin with AA^+

$$AA^+ = (U\Sigma V^T)(V\Sigma^+U^T)$$

Simplify to get:

$$AA^+ = U\Sigma I\Sigma^+U^T$$

Since we have shown before that $\Sigma\Sigma^+$ is an mxm matrix with 1's on the diagonal up to the rth row with zeroes, $\Sigma\Sigma^+$ acts like the identity up to the rth row with all other values brought 0:

$$= U_rU_r^T$$

(e) Previously, we have shown that $AA^+ = U_rU_r^T$. Assuming that A has full row rank, then r = dimension of row space of original matrix = m. This implies that

$$AA^+ = U_mU_m^T$$

And since U is orthonormal,

$$U_mU_m^T = I_m$$

(f) Begin with A^+A

$$AA^+ = (V\Sigma^+U^T)(U\Sigma V^T)$$

Simplify to get:

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

Since we have shown before that $\Sigma^+\Sigma$ is an nxn matrix with 1's on the diagonal up to the rth row with zeroes, $\Sigma^+\Sigma$ acts like the identity up to the rth row of the diagonal with all other values brought 0:

$$= V_r^TV_r$$

(g) Previously, we have shown that $AA^+ = V_r^TV_r$. Assuming that A has full column rank, then r = dimension of column space of original matrix = m. This implies that

$$AA^+ = V_n^TV_n$$

And since V and M are orthonormal ,

$$V_n^TV_n = I_n$$

(h) We can start with the fact that if A is invertible, it has full row and column rank and is a square matrix. Thus, this is true:

$$I = A^+A = AA^+$$

And we can manipulate them to say:

$$I = A^+A \implies A^{-1} = A^+$$

2. Pseudoinverse, projection, and least squares:

   (a) Prove that AA^+ is the projection matrix P onto the column space of A. (Hint: The column space of A is the same as the column space of U_r).
   (b) Prove that A^+A is the projection matrix P onto the row space of A. (Hint: The row space of A^T is the same as the column space of V_r).
   (c) Consider the equation $A\mathbf{x} = \mathbf{b}$. Show that the best least squares solution $\hat{\mathbf{x}}$ to this equation is $\hat{\mathbf{x}} = A^+\mathbf{b}$.

   Note that when we discussed least squares solutions in Chapter 4, we required A to have independent columns. Using the SVD and pseudoinverse, we can now do projections and least squares approximation for any matrix. For a matrix with dependent columns, there are an infinite number of solutions to the equation $A^TA\mathbf{x} = A^T\mathbf{b}$, and $A^+\mathbf{b}$ gives the shortest one (the contribution from $N(A)$ is $\mathbf{0}$).

   To compute the pseudoinverse of A in Julia, use the command "pinv(A)" (using the LinearAlgebra package). See Julia code for HW 11 for examples.

   Solution:

   (a) Projection matrix onto C(A) is defined as

   $$A(A^TA)^{-1}A^T$$

   Since $C(A) = C(U_r)$:

   $$P = U_r(U_r^TU_r)^{-1}U_r^T$$

   We also know that U_r is orthonormal, so:

   $$= U_r(I_r)^{-1}U_r^T$$

   Since the inverse of the identity is itself:

   $$= U_rU_r^T$$

   Which, as we proved before is:

   $$AA^+$$

   (b) Projection matrix onto $C(A^T)$ is defined as

   $$A(A^TA)^{-1}A^T$$

   Since $C(A^T) = C(V_r^T)$:

   $$P = V_r^T(V_rV_r^T)^{-1}(V_r^T)^T$$

   We also know that U_r is orthonormal, so:

   $$= V_r^T(I_r)^{-1}V_r$$

   Since the inverse of the identity is itself:

   $$= V_r^+V_r$$

   Which, as we proved before is:

   $$A^+A$$

(c) The projection of a matrix A onto a vector space V can be defined as p. p can be formed by Pb, where P is the projection matrix onto a given column space. Thus, we know:

$$A\hat{x} = Pb$$

However, since we have shown that AA^+ is a projection onto the column space of A, we can say:

$$A\hat{x} = (AA^+)b$$

Multiplying both sides by A^+, we obtain:

$$A^+A\hat{x} = A^+(AA^+)b$$

Since $A^+A = I_r$:

$$I\hat{x} = IA^+)b$$

Which equals:

$$\hat{x} = A^+b$$

3. Suppose A has (full format) singular value decomposition $A = U\Sigma V^T$, where $\mathbf{u}_i$ is the ith column of U and $\mathbf{v}_i$ is the ith column of V, as usual. Prove that for $1 \leq i \leq r$, $A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i$. What is $A^T\mathbf{u}_i$ when $r < i \leq m$?

   Solution:

   Begin with $A = U\Sigma V^T$. This means that:

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T = V\Sigma U^T$$

   Multiply both sides by U to get:

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

   Since U is orthonormal, we know that:

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

   Since Σ is diagonal, we also know that the columns of $(V\Sigma)$ are $\sigma_i v_i$. We also know that the ith column of $(A^T U)$ can be expressed as $A^T u_i$. Since we know that $A^T U = V\Sigma$, the columns must be equal, so

$$A^T u_i = \sigma_i v_i$$

   When $r < i \leq m$, u_i corresponds to the eigenvalue of 0, and correspond to a vector in the null space of A^T. Thus, $A^T u_i$ is simply the zero vector.

4. In this problem we use PCA/SVD for data visualization. We will use the famous data set of four measurement measurements on three types of iris flowers–see

   `https://en.wikipedia.org/wiki/Iris_flower_data_set`

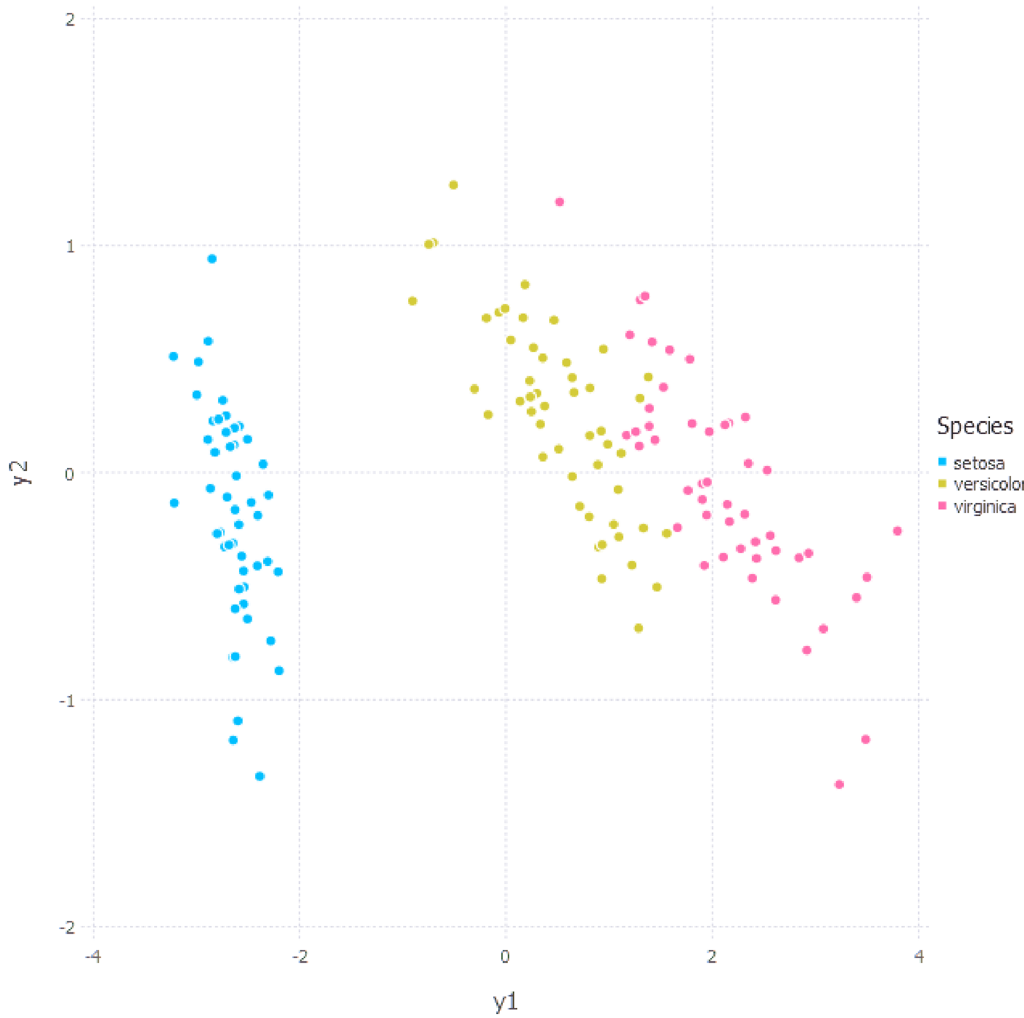
   The first 50 rows are flowers of one species, the next 50 are flowers from a second species, and the last 50 are flowers of a third species. For flower i, the ith row contains measurements of

   $$\mathbf{x}_i^T = (\text{SepalLength}(i), \text{SepalWidth}(i), \text{PetalLength}(i), \text{PetalWidth}(i)),$$

   along with its species in the fifth column. Since $\mathbf{x}_i \in \mathbb{R}^4$ for each i, each flower corresponds to a vector in $\mathbb{R}^4$. We want to visualize this data, so we find the best least-squares projection onto a plane. To do this, you should

   - Import the data set.
   - Center the data to create the 150×4 matrix A.
   - Find the SVD for A, $A = U\Sigma V^T$.
   - Let $\mathbf{y}_1 = A\mathbf{v}_1$. The ith component of the vector $\mathbf{y}_1$ contains the length of the projection of $\mathbf{x}_i$ onto $\mathbf{v}_1$.
   - Let $\mathbf{y}_2 = A\mathbf{v}_2$. The ith component of the vector $\mathbf{y}_2$ contains the length of the projection of $\mathbf{x}_i$ onto $\mathbf{v}_2$.
   - Make a scatter plot of $(\mathbf{y}_1, \mathbf{y}_2)$ where the color of each point corresponds to the species of the flower.
   - What is a typical coefficient for $\mathbf{v}_1$ and $\mathbf{v}_2$ for a given flower of each species?

   The "Julia code for Homework 11" notebook on Canvas shows how to complete each step. Your hw does not need to include your code, but it should include your scatter plot.



Solution:

$Sentosa : -2.3 < Av_1 <, -1.3 < Av_2 < 1$

$veriscolor : -0.5 < Av_1 < 2, -0.6 < Av_2 < 1.2$

$virginica : 0.5 < Av_1 < 4, -1.2 < Av_2 < 1.2$

5. Low rank approximation: In this problem we will find a low-rank approximation for a data matrix. We will use a data set of heptathlon scores for 25 athletes in the heptathlon (7 events) from the 1988 Olympics (25 rows, 7 columns–note the original data matrix has nine columns, where the first column is the athlete's name and the last column is their cumulative score. There are three races in the heptathlon (hurdles, 200m, 800m, columns 1, 4, and 7) so the lower scores for these events are better, while for the rest of the events (high jump, long jump, shot put, javelin) higher scores are better. Similar to the previous problem, you first want to import the data set and center it to get the 25×7 data matrix A. The "Julia code for Homework 11" notebook on Canvas shows how to complete each step.

   Let's see if doing PCA on A tells us anything useful.

   - Find the SVD for A, $A = U\Sigma V^T$.
   - What event explains almost all of the variance (find the element of $\mathbf{v}_1$ that is much larger than all the others).
   - The first few principal components are each close to a standard basis vector (one component dominates. What are these components telling us?

   The example above shows how PCA is sensitive to the units of a data set. To find some more useful information, let's create another matrix B where we normalize each column so it is a unit vector. (Statistically, we are now working with with correlations, not covariances).

   Just to make our results easier to interpret for humans, let's also negate the entries in columns 1, 4, and 7 (note they are still unit vectors!) so that now for every column, a higher number corresponds to a better performance.

   - Can you interpret the first couple of principal components? (Before looking at the data, I expected there might be one component corresponding to track events, one to jumping events, one to throwing events, etc. But it doesn't work out that way. The first principal component seems to correspond to general excellence (or deficiency) in all events except one.)
   - Compute the rank 1 approximation for B. Find the coefficient c_1 so that the first row of B is approximated by $c_1\mathbf{v}_1$.
   - Compute the rank 2 approximation for B. Find the coefficient c_2 so that the first row of B is approximated by $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

   Solution:

   (a) The 800 meter run explains almost all the variance, as it corresponds to the 7th value of $\mathbf{v_1}$, which is much larger than the others

   (b) The fact that the first few principle component vectors appear to be standard basis vectors implies that certain events have more priority than others. The first few principle components are close to e_7, e_6, e_3, which have the highest average values in descending orders. This logically leads to the conclusion that larger numbers(different units) may bias our PCA.

(c) The first principle component is pretty evenly distributed , except for the 6th entry, which is way higher. Since this corresponds to javelin, it indicates that better performance in javelin does not correspond to better performance in other areas. The second principle component has values 1,2,4,7 positive, and columns 3,4,6 negative. Since 1,2,5,7 correspond to hurdles,highjump, long jump and run800, while 3,4,6 correspond to shot, run200 and javelin, it indicates that long jump, hurdles,run800 and high jump require different skill sets. this makes sense, as these are stamina events, while the other three are bursts events.

(d) We can obtain this coefficient by projecting onto the rank 1 approximation and obtaining the first element. This is B1*V[:,1], which turns out to be -0.8412869738489555.

(e) Using the same strategy a before, we obtain -0.253605

6. Generate a perspective drawing with the following specifications:

- Your image plane ia 2×2, containing 1,000,000 square pixels (1000×1000).
- Your camera is 1 unit behind the center of the image plane.

(a) What is your camera basis?

(b) Use the procedure outlined in class to find the "camera coordinates" and pixel locations for the following points (camera at the origin)

$$(1,1,5),(-1,1,5),(-1,-1,5),(1,-1,5),(1,1,8),(-1,1,8),(-1,-1,8),(1,-1,8)$$

(c) Plot the image of these eight points. Note they form the corners of a box. How many sides of the box could you see in this rendering if it were solid and opaque?

(d) Plot the image of the same box in two other locations in space, one where two sides of the box are visible, and one where three sides of the box are visible.

Solution:

(a) $a_1 = \begin{bmatrix} 0.001 \\ 0 \\ 0 \end{bmatrix}, a2 = \begin{bmatrix} 0 \\ 0.001 \\ 0 \end{bmatrix}, a3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, *so* $C = \begin{bmatrix} 0.001 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) let X = matrix of camera coordinates. This is:

$$X = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 5 & 5 & 5 & 5 & 8 & 8 & 8 & 8 \end{bmatrix}$$

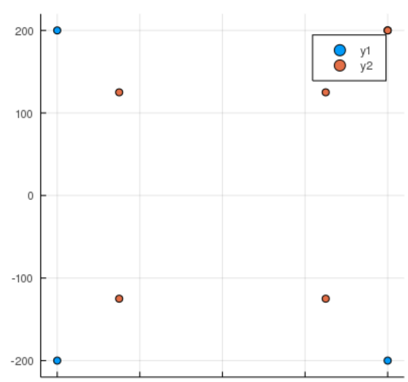
Find camera coordinates with $A^{-1}X$. This is

$$A^{-1}X = \begin{bmatrix} 1000 & -1000 & -1000 & 1000 & 1000 & -1000 & -1000 & 1000 \\ 1000 & 1000 & -1000 & -1000 & 1000 & 1000 & -1000 & -1000 \\ 5 & 5 & 5 & 5 & 8 & 8 & 8 & 8 \end{bmatrix}$$

Now scale by third coordinate to normalize:

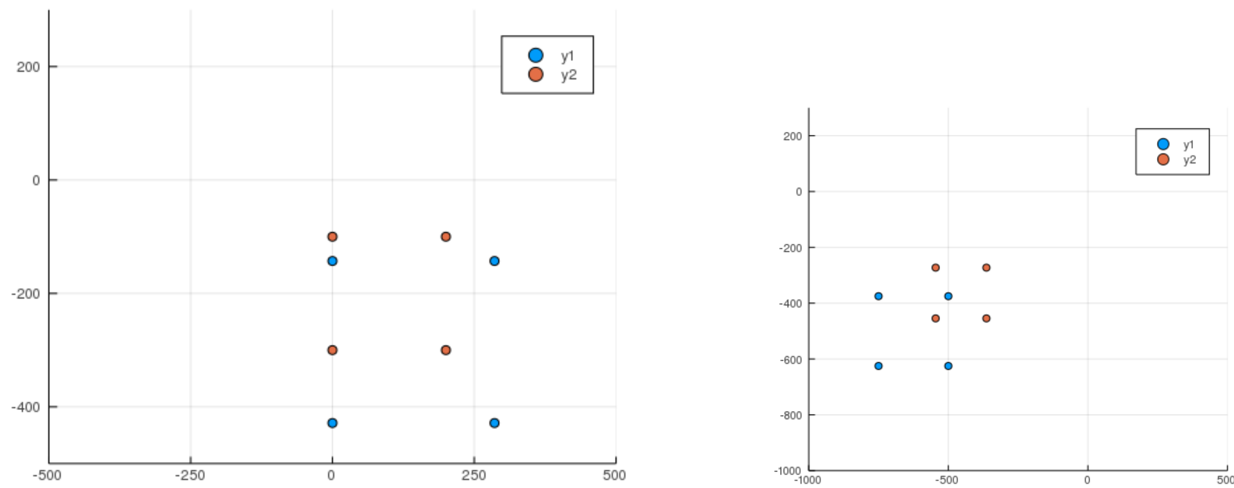
$$\begin{bmatrix} 200 & -200 & -200 & 200 & 125 & -125 & -125 & 125 \\ 200 & 200 & -200 & -200 & 125 & 125 & -125 & -125 \end{bmatrix}$$

(c) Since this is a top view, you would only be able to see one side if it were



opaque.

(d)



y1
y2
200
0
-200
-400
-500
-250
0
250
500
y1
y2
200
0
-200
-400
-600
-800
-1000
-1000
-500
0
500