

MATH661 HW03 - SVD applications

Posted: 09/13/23

Due: 09/20/23, 11:59PM

Tracks 1 & 2: 1. **Track 2:** 2.

This homework marks your first foray into realistic scientific computation by applying concepts from linear algebra, specifically using the singular value decomposition to analyze hurricane Lee, still active at this time.

1 Problem setup

1.1 Image data

Processing of image data through the tools of linear algebra is often encountered, and in this assignment you shall work with satellite images of hurricane Lee of 2023 Season. Open the following folds and execute the code to set up your environment.

- The following commands within a **BASH** shell brings up an animation of hurricane Lee (ImageMagick package must be installed and in the system path).
- The model will use various Julia packages to process image data, carry out linear algebra operations. The following commands build the appropriate Julia environment. The Julia packages are downloaded, compiled and stored in your local Julia library. Note: compiling the Images package takes a long time and it's best to do this within a terminal window. Compiling the packages need only be done once.
- Once compiled, packages are imported into the current environment
- With appropriate packages in place and available, the satellite imagery can be imported into the Julia environment and used to insert images such as the one in Fig. (1).
- Image data must be quantified, i.e., transformed into numbers prior to analysis. In this case matrices of 32-bit floats are obtained from the gray-scale intensity values associated with each pixel. The 720 by 850

pixel images might be too large for machines with limited hardware resources, so adapt the window to obtain reasonable code execution times, Fig. 2.

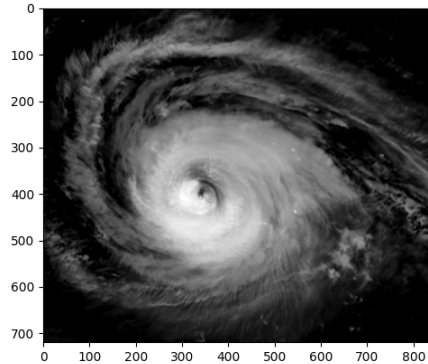


Figure 1. Night-time satellite imagery of hurricane Lee. Superimposed on the familiar overall anti-cyclone pattern are small-scale features (lightning, cloud patterns in the arms). The SVD can be used to distinguish between small and large scale features.

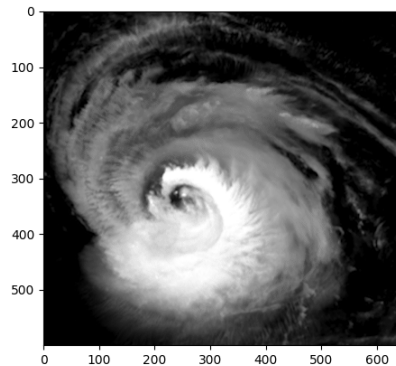


Figure 2. Data window of a time snapshot of hurricane Lee.

1.2 Computing the SVD

The SVD of the array of floats obtained from an image identifies correlations in gray-level intensity between pixel positions, an encoded description of weather physics. Here are the Julia instructions to compute the SVD of one frame of image data. It is instructive to plot the singular values $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots)$, in log coordinates.

```
∴ n=32; A=data[:, :, n]; U,S,V=svd(A);  
∴ clf(); plot(log10.(S), "o"); xlabel(L"mode_number_␣k␣"); ylabel(L"lg(␣sigma_k␣");  
∴ title("Singular_␣values_␣of_␣hurricane_␣Lee_␣image"); grid("on");  
∴ hwdir=homedir()*"/courses/MATH661/homework/H03"; savefig(hwdir*"/H03Fig03a.eps");
```

```

∴ clf(); plot(log10.(S[1:100]),"o"); xlabel(L"mode_number_$$$");
ylabel(L"lg(\sigma_k)");
∴ title("Singular_values_of_hurricane_Lee_image"); grid("on");
∴ savefig(hwdir*"/H03Fig03b.eps");
∴

```

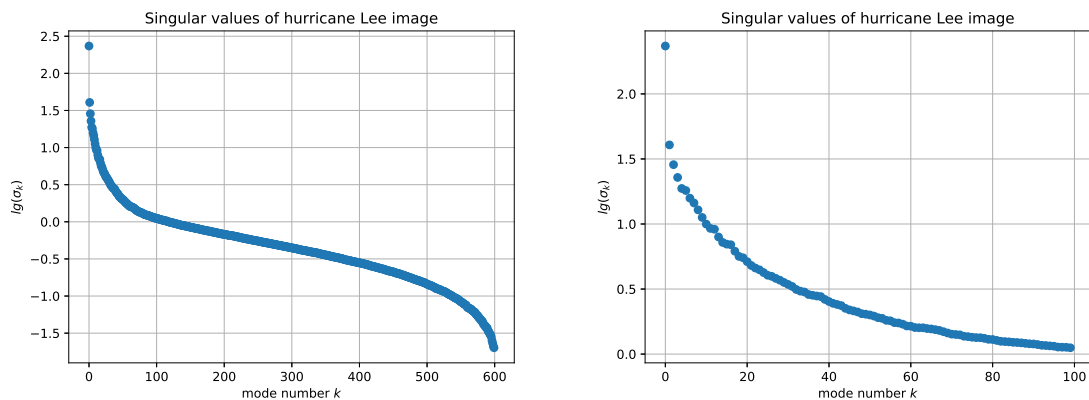


Figure 3. The singular values of a hurricane satellite image rapidly decay from $\sim 10^{2.5}$ to $\sim 10^0$, a two order-of-magnitude decrease over the first 100 modes (singular vectors). This indicates considerable data compression is possible.

- Define a function `rsvd` to sum the rank-one updates from p to q from an SVD,

$$A = U \Sigma V^T \cong B = \sum_{k=p}^q \sigma_j \mathbf{u}_j \mathbf{v}_j^T.$$

The sum from $p = 1$ to q contains the first q dominant correlations, identified as large scale weather patterns. The sum from $p = 101$ to $q = 140$ can be identified as small scale patterns.

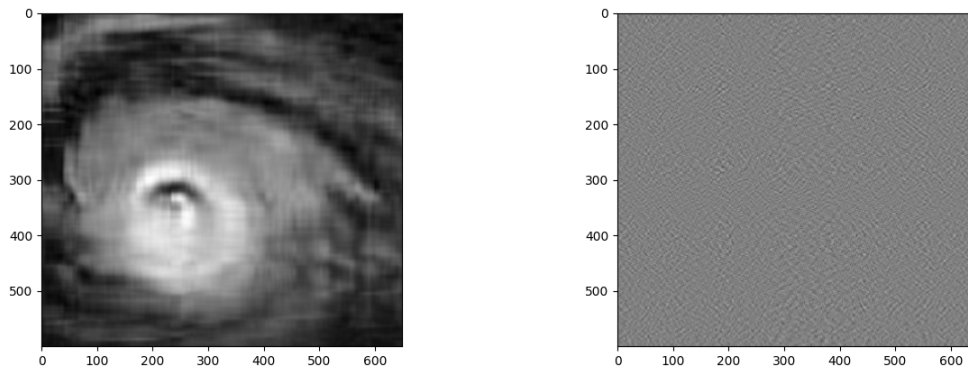


Figure 4. (Left) Reconstruction of hurricane Lee image from first $q = 12$ modes, showing large scale patterns. (Right) Correlated small-scale patterns obtained by sum of modes $p = 101$ to $q = 140$.

2 Common problems (both tracks)

1. Consider large-scale weather patterns B_k, B_{k-1}, B_{k-2} obtained from the p most significant modes from frames $k, k-1, k-2$ (i.e., different times in the past). Assuming a constant rate of change leads to the prediction

$$P_k = B_{k-1} + (B_{k-1} - B_{k-2}) \quad (3)$$

for the known large-scale weather B_k . The prediction error is $\varepsilon_{k,1}(p) = \|P_k - B_k\|_F / \|B_k\|_F$. Present a plot of the prediction error for various values of p over the recorded hurricane data. Analyze your results to answer the question: “can overall storm evolution be predicted by linear extrapolation of past data?”

- Solution.**
- Define a function to construct prediction at time k by linear extrapolation of modes from p to q at times $k-1, k-2$
 - Define a function to compare the coarse grained prediction and data
 - Define a function to plot the prediction error over the available data

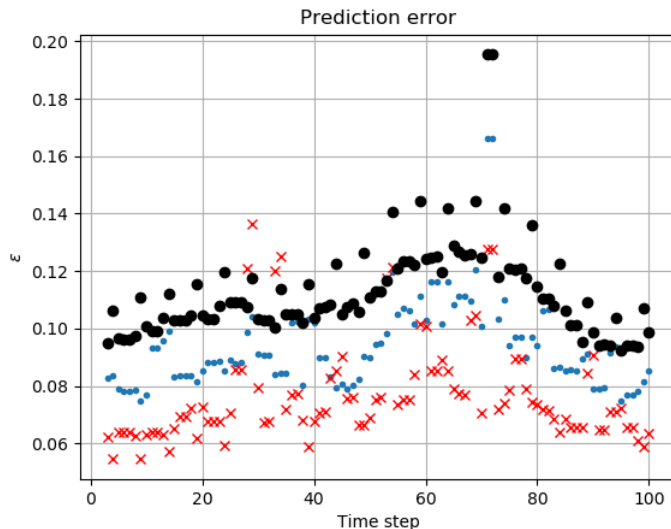


Figure 5. Linear extrapolation error, unprocessed images: (\times) modes $p = 1$ to $q = 5$; (\bullet) modes $p = 1$ to $q = 10$; (\bullet) modes $p = 1$ to $q = 20$.

Relative error (Fig. 5) between prediction and observed data increases as with increase in the number of SVD modes (Fig. 6). This might seem surprising at first glance, but the comparison is done on raw data. In almost all real-world applications additional pre/post-processing of data is needed to:

- ensure identical illumination in the two images;
- align and rotate the images to match overall features (aka “image registration”)
- determine whether pixel-by-pixel comparison is meaningful as opposed to derived measures: total cloud cover, position of center, etc.

Bonus point project (4 points). Carry out the above operations. Of particular relevance is the recognition of overall structure in data, in this case the spiral pattern of the anti-cyclone. Think of a way to use this knowledge of the

expected overall pattern to better compare predictions to data.

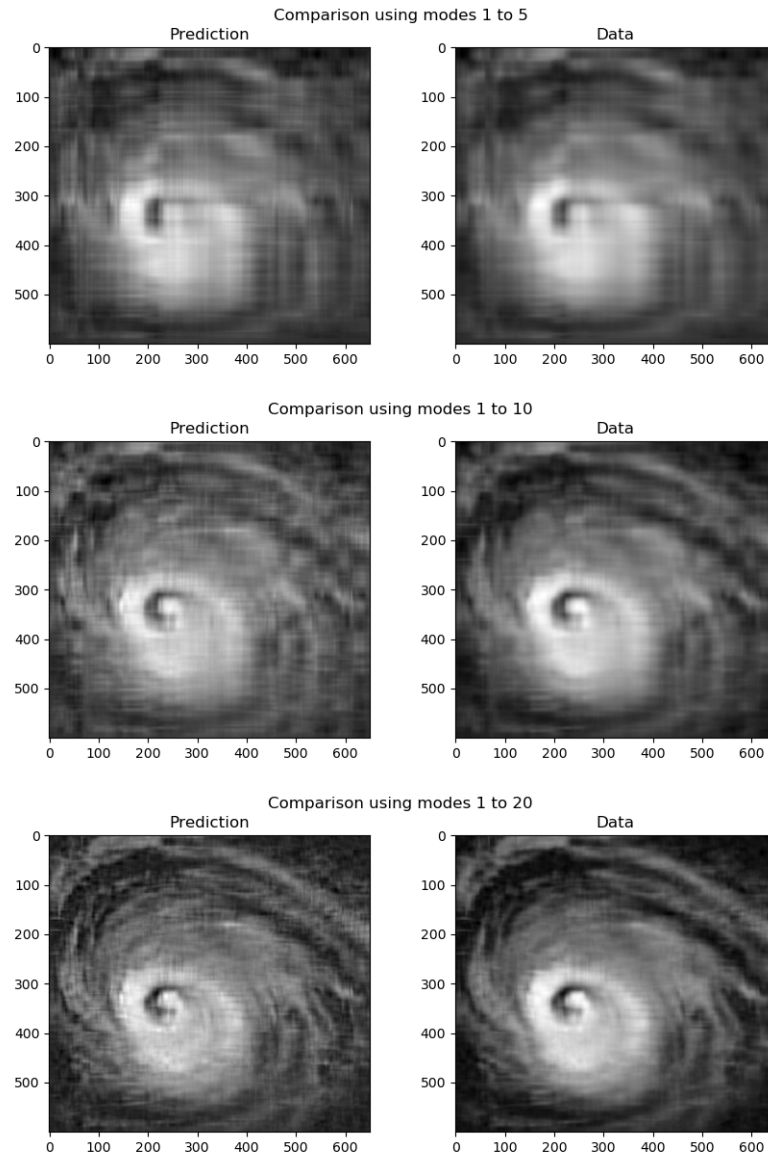


Figure 6. Comparison of coarse-grained prediction with data for 5, 10, 20 modes from top to bottom.

2. **Repeat** the above construction of an error plot for local weather patterns C_k, C_{k-1}, C_{k-2} for lesser significance modes from p to q , that lead to prediction

$$Q_k = C_{k-1} + (C_{k-1} - C_{k-2}).$$

Analyze your results to answer the question: “are small-scale storm features predictable?”

Solution. The key word in the problem statement is “repeat”. Defining purpose-specific functions is not only good coding practice, but saves time in studying similar problems. Here, the previously defined functions are simply

invoked with new parameters.

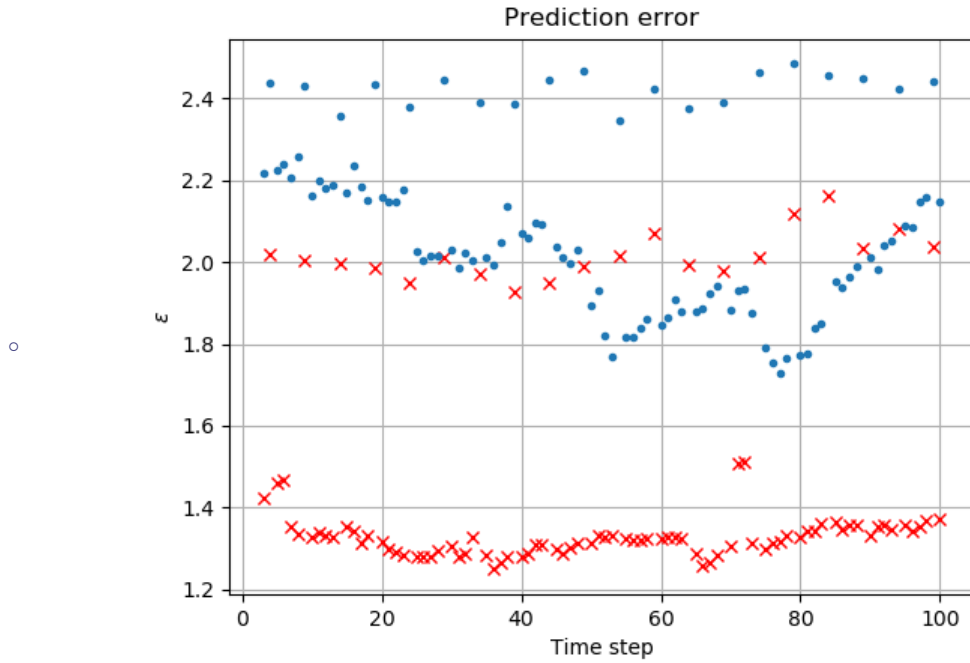


Figure 7. Linear extrapolation error, unprocessed images: (×) modes $p=30$ to $q=60$; (•) modes $p=60$ to $q=90$.

The $\mathcal{O}(1)$ relative error (Fig. 7) would at first glance indicate no predictive capability of the linear extrapolation of the fine scale structure. However the error definition is inappropriate:

- the error measure is global while the modes reflect structure at smaller spatial scales
- illumination and registration has not been performed.

As in the previous case, several additional steps have to be carried out on the fine scale data in order to determine whether linear extrapolation is an efficient predictor. In particular, the smaller spatial extent has to be taken into account. A typical processing sequence would:

- Ensure equal illumination and registration
- Define a window of smaller size, and construct the mean and variance observed when translating the window over the entire image. This would be a data-extracted model of small-scale features of the hurricane
- Use observed mean and variance to construct extrapolated small-scale features.

3. Repeat the above for combined large (B) and small scale (C) weather patterns. Experiment with different weights u, v in the linear combination $uB + vC$, $u + v = 1$.

Solution.

- Define a function to combine both large and small scales into a weighted average
- Define a function to compare the combined large/small-scale prediction and data

- Define a function to plot the prediction error over the available data

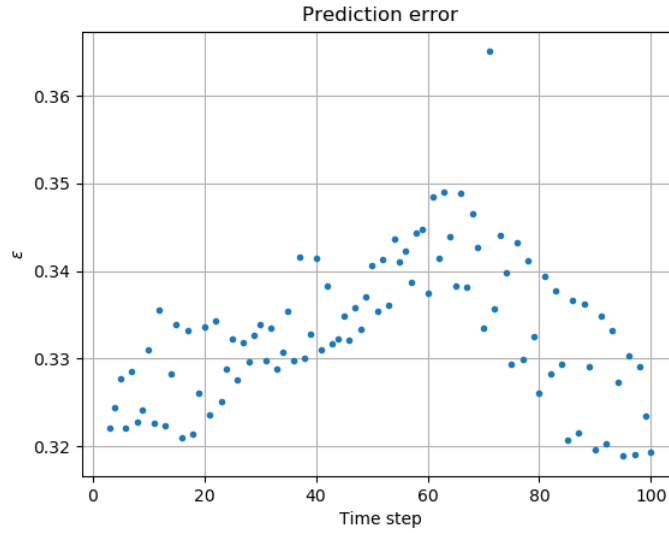


Figure 8. Linear extrapolation error using large and small modes

4. Formula (3) expresses a degree-one in time t prediction

$$B(t) = B_{k-1} + (t - k + 1)(B_{k-1} - B_{k-2}),$$

evaluated at $t = k$. Construct a quadratic prediction based upon data $B_{k-1}, B_{k-2}, B_{k-3}$, and compare with degree-one prediction in question 1. Recall that a quadratic can be constructed from knowledge of three points. Again, experiment with various values of k, p .

Solution. Consider $y(t)$ with known values $y_i = y(i)$ at $i = 0, 1, 2$. The quadratic passing through these three points is

$$y(t) = \frac{1}{2}(t-1)(t-2)y_0 - t(t-2)y_1 + \frac{1}{2}t(t-1)y_2,$$

and predicts value

$$y(3) = y_0 - 3y_1 + 3y_2,$$

hence quadratic predictions are given by

$$B_k = B_{k-3} - 3B_{k-2} + 3B_{k-1}.$$

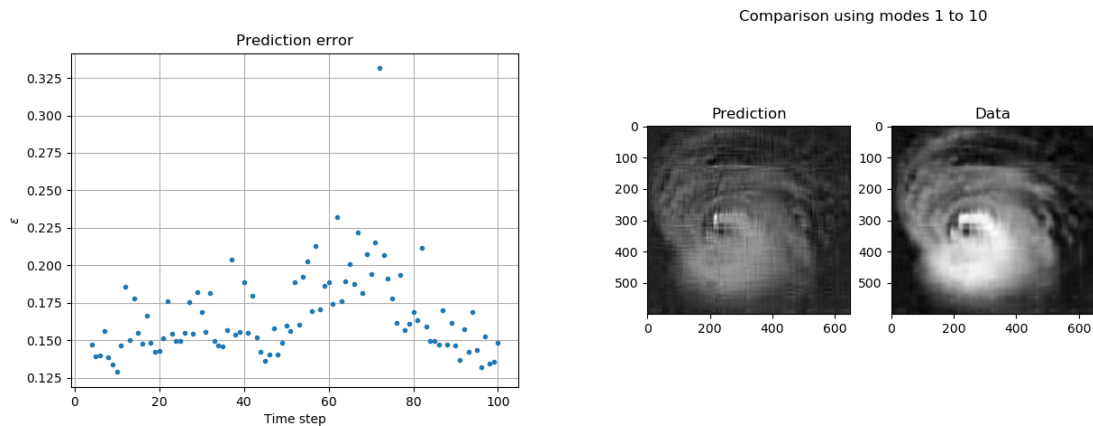


Figure 9. (Left) Errors of large scale quadratic prediction; (Right) Comparison.

Errors for quadratic extrapolation (Fig. 9) are higher than those for linear extrapolation (Fig. 5). In addition to the errors discussed previously (illumination, registration), higher order extrapolation generally exhibits higher error (to be discussed during presentation of polynomial interpolation).

3 Track 2 questions

Use the SVD to investigate how familiar properties for $a \in \mathbb{R}$ might extend to matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$. For example, on the real axis adding $-a$ to a results in the null element, $a - a = 0$, so we say a is a distance $|a|$ from zero. This easily generalizes to \mathbb{R}^m , and the minimal modification of $\mathbf{a} \in \mathbb{R}^m$ to obtain zero is $-\mathbf{a}$, $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$, and \mathbf{a} is distance $\|\mathbf{a}\|$ away from zero. Consider now matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$, with $\text{rank}(\mathbf{A}) = m$. How far is this matrix from “zero”? By “zero” we shall understand a singular matrix with $\text{rank}(\mathbf{A}) < m$.

All real numbers $x \in \mathbb{R}$ can be obtained as the limit of a sequence of rationals $\{p_n/q_n\}$, $p_n, q_n \in \mathbb{Z}$. We say that the rationals \mathbb{Q} is a dense subset of \mathbb{R} . What about matrix spaces?

1. Find \mathbf{X} of minimal norm such that $\mathbf{A} + \mathbf{X}$ is singular. State how far \mathbf{A} is from “zero”?

Solution. Seek insight from $m = 1$ for which the problem is to find x of minimal norm such that $a + x = 0$. From

$$0 = |a + x| \geq |a| - |x| \Rightarrow |x| \geq |a| \quad \text{and} \quad 0 = |x + a| \geq |x| - |a| \Rightarrow |x| \leq |a|$$

deduce that $|x| = |a|$ and the solution is $x = -a$. Recall that $\|\mathbf{A}\|$ (the induced 2-norm in the context of an SVD), the maximal amplification factor along any direction within \mathbb{R}^m , becomes $|a|$, interpreted as the amplification factor when $m = 1$. What changes when $m > 1$? There are now different amplification factors of \mathbf{A} along different directions. The maximum amplification factor (i.e., the matrix norm) is given by the largest singular value

$$\sigma_1 = \|\mathbf{A}\| = \sup_{\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\|, \quad \text{and} \quad \|\mathbf{A}\mathbf{u}\| \leq \sigma_1 \|\mathbf{u}\|.$$

There is now also a minimal amplification factor, the smallest singular value

$$\sigma_m = \inf_{\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\|, \quad \text{and} \quad \|\mathbf{A}\mathbf{u}\| \geq \sigma_m \|\mathbf{u}\|.$$

The matrix \mathbf{A} is singular if $\sigma_m = 0$, stating that along at least one direction \mathbf{v} the scaling factor is null, $\mathbf{A}\mathbf{v} = \mathbf{0}$, in which case $\mathbf{X} = \mathbf{0}$ is of minimal norm and trivially satisfies the problem conditions. Assume now that \mathbf{A} is not singular, $\sigma_m > 0$, and investigate the *smallest* amplification factor of $\mathbf{A} + \mathbf{X}$, seeking a lower bound for $\|(\mathbf{A} + \mathbf{X})\mathbf{u}\|$ along some arbitrary direction \mathbf{u} ,

$$\|(\mathbf{A} + \mathbf{X})\mathbf{u}\| = \|\mathbf{A}\mathbf{u} + \mathbf{X}\mathbf{u}\| \geq \|\mathbf{A}\mathbf{u}\| - \|\mathbf{X}\mathbf{u}\| \geq (\sigma_m - \|\mathbf{X}\|)\|\mathbf{u}\|.$$

The above states that if $\|\mathbf{X}\| < \sigma_m$ then $\|(\mathbf{A} + \mathbf{X})\mathbf{u}\| > 0$ for $\|\mathbf{u}\| = 1$ and $\mathbf{A} + \mathbf{X}$ is not singular. Hence, in order for $\mathbf{A} + \mathbf{X}$ to be singular the condition $\|\mathbf{X}\| \geq \sigma_m$ must hold. Use the SVD of $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and construct

$$\mathbf{X} = \mathbf{U}\text{diag}(0, \dots, 0, -\sigma_m)\mathbf{V}^T$$

that makes the sum $\mathbf{A} + \mathbf{X}$ singular, and $\|\mathbf{X}\| = \sigma_m$, the minimal allowed norm.

2. Prove that any matrix in $\mathbb{R}^{m \times m}$ is the limit of a sequence of matrices of full rank, i.e., the set of full-rank matrices is a dense subset of $\mathbb{R}^{m \times m}$.

Solution. The SVD of \mathbf{A} of rank r is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ with $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$. Define sequences $\{s_n^{(j)}\}_{n \in \mathbb{N}}$ such that

$$\lim_{n \rightarrow \infty} s_n^{(j)} = \sigma_j, \quad \text{e.g.}$$

$$s_n^{(j)} = \frac{n+1}{n} \sigma_j \quad \text{for } j \leq r, \quad s_n^{(j)} = \frac{n+1}{n^2} \quad \text{for } j > r.$$

Then $\mathbf{A}_n = \mathbf{U}\mathbf{S}_n\mathbf{V}^T$ with $\mathbf{S}_n = \text{diag}(s_n^{(1)}, s_n^{(2)}, \dots, s_n^{(m)})$ are of full rank and $\lim_{n \rightarrow \infty} \mathbf{A}_n = \mathbf{A}$.