Problem 7.1

(a) Let λ be an eigenvalue of M, and $x \in \mathbb{R}^n$ an associated eigenvector of norm 1. Compute:

$$x^T M x = x^T(\lambda x) = \lambda x^T x = \lambda.$$

Since $x \neq 0$, we know that $x^T M x > 0$ which gives that $\lambda > 0$: the eigenvalues of M are strictly positive.

In particular, we get that 0 is not an eigenvalue of M hence $\mathrm{Ker}(M) = \{0\}$. This gives that M is invertible.

(b) Let λ_{min} be the smallest eigenvalue of M. We have then:

This can be osed without proof. For completeness I give a proof below.

$$\boxed{\lambda_{min} = \min_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{x^T M x}{\|x\|^2}}.$$

Indeed, notice that $-\lambda_{min}$ is the largest eigenvalue of $-M$, hence we know from the lecture that:

$$-\lambda_{min} = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{-x^T M x}{\|x\|^2} = -\min_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{x^T M x}{\|x\|^2}.$$

Let now $\alpha = |\lambda_{min}| + 1$. Let $x \in \mathbb{R}^n \setminus \{0\}$

$$x^T(M + \alpha \mathrm{Id}_n) x = x^T M x + \alpha \|x\|^2$$
$$= \|x\|^2 \left(\frac{x^T M x}{\|x\|^2} + \alpha \right)$$

$$\geqslant \|x\|^2\left(\lambda_{min} + |\lambda_{min}| + 1\right) > 0$$

Which proves that $M + \alpha Id_n$ is positive definite

Problem 7.2.

a) Let $A = \begin{pmatrix} — a_1^T — \\ \\ — a_n^T — \end{pmatrix} \in \mathbb{R}^{n \times d}$

Let $(v_1 \dots v_d)$ be an orthonormal basis of $\mathbb{R}^d$ consisting of eigenvectors of A^TA, associated with the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_d$. (since A^TA is symmetric, the spectral theorem guarantees that such vectors exists).

For $i \in \{1, \dots n\}$, b_i is given by:

$$b_i = \begin{pmatrix} \langle v_1, a_i \rangle \\ \vdots \\ \langle v_k, a_i \rangle \end{pmatrix} \in \mathbb{R}^k$$

Compute:

$$b_1 + \dots + b_n = \begin{pmatrix} \langle v_1, \overbrace{a_1 + \dots + a_n}^{=0} \rangle \\ \vdots \\ \langle v_k, \underbrace{a_1 + \dots + a_n}_{=0} \rangle \end{pmatrix}$$

because the a_i are centered.

$$= 0.$$

b) Let $i, j \in \{1, \dots n\}$

$$\|b_i - b_j\|^2 = \langle v_1, a_i - a_j\rangle^2 + \dots + \langle v_k, a_i - a_j\rangle^2$$

(because $k \leq d$)

$$\leq \underbrace{\langle v_1, a_i - a_j\rangle^2 + \dots + \langle v_d, a_i - a_j\rangle^2}$$

$= \|a_i - a_j\|^2$ because $(v_1, \dots v_d)$ is an orthonormal basis of $\mathbb{R}^d$

c) $$f^{(i)} = \begin{pmatrix} \langle a_1, v_i\rangle \\ \vdots \\ \langle a_n, v_i\rangle \end{pmatrix} = A v_i$$

Hence $$\langle f^{(j)}, f^{(i)}\rangle = v_j^T A^T A v_i = \lambda_i \langle v_j, v_i\rangle$$

Since v_i is an eigenvector of $A^T A$ associated with the eigenvalue λ_i.

Consequently, when $i \neq j$, $\langle f^{(j)}, f^{(i)}\rangle = 0$ because $\langle v_j, v_i\rangle = 0$ since $(v_1 \_ v_d)$ is orthogonal.

Problem 7.3.

a) We have:

$$U\Sigma = \begin{array}{c} \begin{array}{cc} \overbrace{\hphantom{u_1 \cdots u_r}}^{r} & \overbrace{\hphantom{\cdots u_n}}^{n-r} \end{array} \\ \begin{pmatrix} | & & | & & | \\ u_1 & \cdots & u_r & \cdots & u_n \\ | & & | & & | \end{pmatrix} \end{array} \begin{array}{c} \begin{array}{c} r \\ n-r \end{array} \end{array} \begin{pmatrix} \tilde{\Sigma} & (0) \\ (0) & (0) \end{pmatrix}_{\begin{array}{cc} r & m-r \end{array}}$$

$$= \tilde{U} \left(\underbrace{\tilde{\Sigma}}_{r} \;\middle|\; \underbrace{(0)}_{m-r} \right)$$

Similarly:

$$\left(\tilde{\Sigma} \,|\, (0)\right) V^T = \left(\overbrace{\tilde{\Sigma}}^{r} \;\middle|\; \overbrace{(0)}^{m-r} \right) \begin{array}{c} r \\ m-r \end{array} \begin{pmatrix} \text{———} & v_1^T & \text{———} \\ \text{———} & v_r^T & \text{———} \\ \text{———} & v_m^T & \text{———} \end{pmatrix}$$

$$= \tilde{\Sigma} \tilde{V}^T.$$

Conclusion: $A = U\Sigma V^T = \tilde{U}\left(\tilde{\Sigma} \,|\, (0)\right) V^T$

$$= \underline{\tilde{U} \tilde{\Sigma} \tilde{V}^T}$$

b) Since $\text{rank}(A) = r$, the rank-nullity theorem gives that $\dim \text{Ker}(A) = m - r$

For $i \in \{r+1, \dots m\}$,

$$\tilde{V}^T v_i = \begin{pmatrix} \langle v_1, v_i \rangle \\ \vdots \\ \langle v_r, v_i \rangle \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$ because $(v_1 \dots v_m)$ is orthonormal.

Hence for all $i \in \{r+1, \dots m\}$, $A v_i = \tilde{U} \tilde{\Sigma} \tilde{V}^T v_i = 0$
$v_i \in \text{Ker}(A)$.

Conclusion: $(v_{r+1}, \dots, v_m)$ is an orthonormal family of vectors of $\text{Ker}(A)$. Since $\dim \text{Ker}(A) = m - r$, this is an orthonormal basis of $\text{Ker}(A)$.

For $i \in \{1, \dots r\}$, compute:

$$A\left(\frac{v_i}{\sigma_i}\right) = \frac{1}{\sigma_i} \tilde{U} \tilde{\Sigma} \tilde{V}^T v_i$$

where $\tilde{e}_1 \dots \tilde{e}_r$ denotes the canonical basis of $\mathbb{R}^r$

$$\left(\begin{array}{c}\text{sine } v_1 \dots v_r \\ \text{orthonormal}\end{array}\right) = \frac{1}{\sigma_i} \tilde{U} \tilde{\Sigma} \tilde{e}_i$$

$$= \frac{1}{\sigma_i} \tilde{U} \sigma_i e_i$$

$$= u_i$$

Hence for all $i \in \{1, \dots r\}$, $u_i \in \text{Im}(A)$

Conclusion: $(u_1, \dots u_r)$ is an orthonormal family of $\mathrm{Im}(A)$. Since $\dim \mathrm{Im}(A) = r$, it is an orthonormal basis of $\mathrm{Im}(A)$.

In [1]:

```
%matplotlib inline
import matplotlib
import numpy as np
import matplotlib.pyplot as plot
```

In [2]:

```
# Load the data matrix
A = np.loadtxt('mysterious_data.txt')
n,d = A.shape
print(f'The matrix A contains {n} points in dimension {d}')
```

```
The matrix A contains 6344 points in dimension 1000
```

Each row of A corresponds to a datapoint.

In [3]:

```
# We are going to invetigate the dataset using PCA

# First, we center the dataset
A = A - A.mean(axis=0)
```

In [4]:

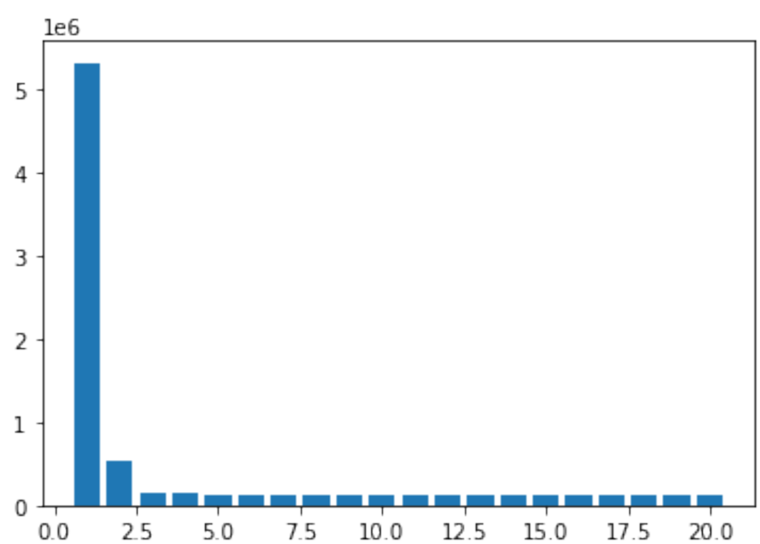
```
# Then, we compute the covariance matrix (up to a factor 1/n)
S = A.T @ A

# and its eigenvalues/eigenvectors:
lamb, V = np.linalg.eigh(S)

# The eigenvalues of S are ranked in the ascending order,
# and the corresponding eigenvectors are the columns of V.
# Since we are interested in the largest eigenvalues, we
# reverse the order:
lamb = lamb[::-1]
```

In [6]:

```
# We plot the first m largest eigenvalues
m=20
plot.bar(np.array(range(m))+1, lamb[:m])
```

Out[6]:

```
<BarContainer object of 20 artists>
```



In [11]:

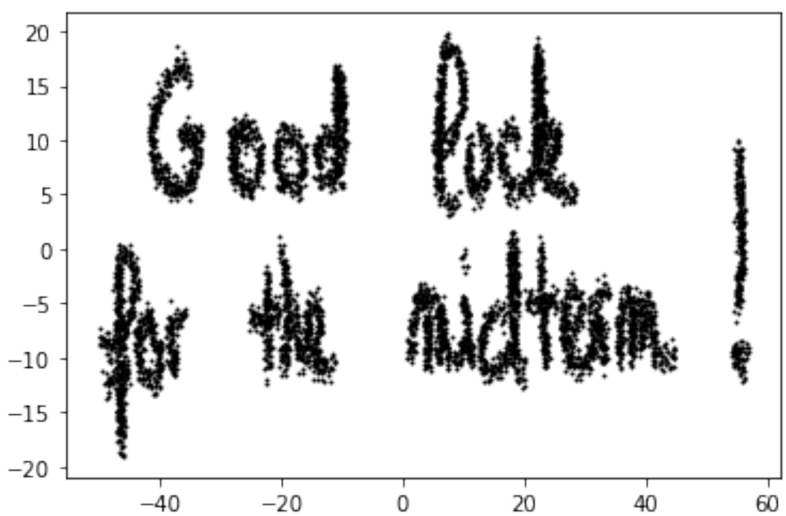
```python
# On the plot, we see that the first two eigenvalues are significantly
# larger than all the others.
# This means that there are two directions in which
# the dataset has a much larger variance than the others.
# This encourages us to visualise the dataset using only the two first
# principal components.

# The first principal component of each point is
x = A @ V[:,-1]
# and the second is
y = A @ V[:,-2]
```

In [14]:

```python
plot.scatter(-x,-y,s=1 ,color='black')

# Oh, such a nice message!
# I had to flip x and y to get it in the right position.
# This is not a problem because principal components
# are only defined up to a global sign flip.
# Indeed if v1, ..., vd are an orthonormal basis of eigenvectors of S
# then so is -v1,v2, ...,vd.
```

Out[14]:

```
<matplotlib.collections.PathCollection at 0x11d09d400>
```



In [ ]:

This gives $x^T M x \geqslant \lambda_n \|x\|^2$.

Let $\alpha = |\lambda_n| + 1$.

$$x^T (M + \alpha Id_n) x = x^T M x + \alpha x^T x$$
$$\geqslant \lambda_n \|x\|^2 + (|\lambda_n| + 1) \|x\|^2$$
$$\geqslant \|x\|^2 > 0 \text{ for } x \in \mathbb{R}^n \setminus \{0\}.$$

<u>Conclusion</u> : $M + \alpha Id_n$ is positive definite.

7.4. Let $v_1 \dots v_n$ be an orthonormal family of eigenvectors of M, associated respectively with $\lambda_1, \dots \lambda_n$.

Let $P = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix}$ and $D = \mathrm{Diag}(\lambda_1 \dots \lambda_n)$.

We know that $M = PDP^T$.

Let $d \leqslant n$.
Let $U \in \mathbb{R}^{n \times d}$ such that $U^T U = Id_d$.
Define $R = P^T U$. and notice that $R^T R = U^T P P^T U$
$$= U^T Id_n U = Id_d$$

$$Tr(U^T M U) = Tr(U^T P D P^T U)$$
$$= Tr(R^T D R)$$
$$= \sum_{i=1}^{d} (R^T D R)_{i,i}$$

$$(R^T D R)_{i,i} = \sum_{j=1}^{n} (R^T)_{i,j} (DR)_{j,i}$$

$$= \sum_{j=1}^{n} R_{j,i} \sum_{k=1}^{n} D_{j,k} R_{k,i}$$

$$= \sum_{j=1}^{n} R_{j,i} \lambda_j R_{j,i} \quad \text{because } D_{j,k} = \begin{cases} \lambda_j & \text{if } k=j \\ 0 & \text{otherwise.} \end{cases}$$

$$= \sum_{j=1}^{n} \lambda_j R_{j,i}^2 .$$

Hence $Tr(U^T M U) = \sum_{i=1}^{d} \sum_{j=1}^{n} \lambda_j R_{j,i}^2$

$$= \sum_{j=1}^{n} \lambda_j . \underbrace{\left(\sum_{i=1}^{d} R_{j,i}^2 \right)}_{\overset{def}{=} w_j}$$

- We have $\sum_{j=1}^{n} w_j = d$. Indeed:

$$\sum_{j=1}^{n} w_j = \sum_{j=1}^{n} \sum_{i=1}^{d} R_{j,i}^2 = Tr(R^T R) = Tr(Id_d) = d.$$

- For all j, $0 \leq w_j \leq 1$. Indeed:

$$0 \leq w_j = e_j^T R R^T e_j = \langle e_j, \underbrace{RR^T}_{\text{orth. projection matrix}} e_j \rangle \leq \|e_j\| \, \|RR^T e_j\| \leq 1.$$

Hence we have n weights $w_1 \ldots w_n$ that sum to d and that are all between 0 and 1.

Hence $\sum_{j=1}^{n} \lambda_j w_j \leq \sum_{j=1}^{d} \lambda_j$ because $\lambda_1 \geq \ldots \geq \lambda_n$.

We get that $Tr(U^T M U) \leq \sum_{j=1}^{d} \lambda_j$.

Since U was arbitrarily chosen, we get that

$$\max_{\substack{U \in \mathbb{R}^{n\times d} \\ U^T U = Id_d}} \operatorname{Tr}(U^T M U) \leq \sum_{j=1}^{d} \lambda_j .$$

Let us now prove the converse inequality.
Let $U = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{pmatrix} \in \mathbb{R}^{n\times d}$.

$v_1 \dots v_d$ are orthonormal, hence $U^T U = Id_d$.

Compute $P^T U = \begin{pmatrix} — v_1 — \\ \vdots \\ — v_n — \end{pmatrix} \begin{pmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{pmatrix} = \begin{pmatrix} 1 & & (0) \\ & \ddots & \\ (0) & & 1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$

Hence $\operatorname{Tr}(U^T M U) = \operatorname{Tr}((P^T U)^T D \, P^T U)$

$$= \operatorname{Tr}\left(\underbrace{\left(\begin{array}{ccc|c} 1 & & (0) & \\ & \ddots & & (0) \\ (0) & & 1 & \end{array}\right)}_{d} \cdot \begin{pmatrix} \lambda_1 & & (0) \\ & \ddots & \\ (0) & & \lambda_n \end{pmatrix} \cdot \begin{array}{c} \left(\begin{array}{ccc} 1 & & (0) \\ & \ddots & \\ (0) & & 1 \\ \hline & (0) & \end{array}\right) \end{array} \begin{array}{c} \updownarrow d \\ \\ \end{array} \right)$$

$$= \sum_{j=1}^{d} \lambda_j .$$ which proves the converse equality.