DS-GA 1014 - Homework 7

Eric Niblock

October 23rd, 2020

1. **(2 points). We say that a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive definite if for all non-zero $x \in \mathbb{R}^n$,**

$$\vec{\mathbf{x}}^T M \vec{\mathbf{x}} > 0$$

If a matrix M is positive definite, then M is also positive semi-definite, but the converse is not true. One of the goals of this problem is to prove one of the implications of *Proposition 1.2* of the notes (Lecture 7). You are of course not allowed to use this proposition to solve this problem.

(a) **Let $M \in \mathbb{R}^{n \times n}$ be a positive definite matrix. Show that its eigenvalues are all strictly positive and that M is invertible.**

We begin by showing that all of the eigenvalues of matrix M are positive. For any eigenvalue-eigenvector pair, we have the following,

$$M\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}} \tag{1}$$

Then we can observe the following implications. For any eigenvector $\vec{\mathbf{v}}$,

$$\vec{\mathbf{v}}^T M \vec{\mathbf{v}} > 0 \tag{2}$$
$$\vec{\mathbf{v}}^T \lambda \vec{\mathbf{v}} > 0 \tag{3}$$
$$\lambda ||\vec{\mathbf{v}}||^2 > 0 \tag{4}$$

By the statement of the problem, we know that $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$, and therefore $||\vec{\mathbf{v}}||^2 > 0$. In order to maintain this equality, it follows that $\lambda > 0$. This statement holds for any eigenvalue-eigenvector pair of M, meaning that if $M \in \mathbb{R}^{n \times n}$ is a positive definite matrix its eigenvalues are all strictly positive. Again, given the equation from (1), and the fact that all $\lambda > 0$ that are associated to M, we have that there is no $\lambda = 0$. Furthermore, this implies that there is no nontrivial $\overrightarrow{\mathbf{v}}$ such that $M\overrightarrow{\mathbf{v}} = 0$. The kernel of M is therefore empty, which implies that the matrix M is invertible.

(b) Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Show that there exists $\alpha > 0$ such that the matrix $M + \alpha Id_n$ is positive definite.

First, we set out to show that if all of the eigenvalues of a symmetric matrix M are positive, then M must be a positive definite matrix. Note the Spectral Theorem,

> *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then there exists an orthogonal matrix P and a diagonal matrix D of sizes $n \times n$, such that,*
>
> $$A = PDP^T$$
>
> *[Prop. 1]*

Therefore, it stands to reason that we can represent M as $M = PDP^T$. It follows readily that,

$$\overrightarrow{\mathbf{x}}^T M \overrightarrow{\mathbf{x}} = \overrightarrow{\mathbf{x}}^T PDP^T \overrightarrow{\mathbf{x}} = \overrightarrow{\mathbf{x}}^T PD(\overrightarrow{\mathbf{x}}^T P)^T \tag{5}$$

Now since D is a diagonal matrix populated with the eigenvalues of M (by Spectral Theory), and $\overrightarrow{\mathbf{x}}^T P \in \mathbb{R}^{1\times n}$, we may as well write $\overrightarrow{\mathbf{x}}^T P$ as a vector $\overrightarrow{\mathbf{y}}^T$. Then,

$$\overrightarrow{\mathbf{x}}^T PD(\overrightarrow{\mathbf{x}}^T P)^T = \overrightarrow{\mathbf{y}}^T D \overrightarrow{\mathbf{y}} = \lambda_1 y_1^2 + ... + \lambda_n y_n^2 \tag{6}$$

Then, by the assumption that all of the eigenvalues of M are positive, we are left with,

$$\overrightarrow{\mathbf{x}}^T M \overrightarrow{\mathbf{x}} = \lambda_1 y_1^2 + ... + \lambda_n y_n^2 > 0 \tag{7}$$

This proves that given a symmetric matrix M, where all of the eigenvalues of M are positive, M must be positive-definite.

We know that M has a set of real eigenvalues because it is symmetric. Furthermore, it stands to reason that at least one of the eigenvalues associated with M is negative, since if all of the eigenvalues of M were positive, it would be a positive definite matrix. Suppose that we find the smallest eigenvector of M such that,

$$M\overrightarrow{\mathbf{v}} = \lambda_{min}\overrightarrow{\mathbf{v}} \tag{8}$$

It stands to reason that we must shift λ_{min} such that $\lambda_{min} > 0$. If we take $\alpha > |\lambda_{min}|$, then we have,

$$(A + \alpha Id_n)\overrightarrow{\mathbf{v}} = (\lambda_{min} + \alpha)\overrightarrow{\mathbf{v}} \tag{9}$$

Where $\lambda_{min} + \alpha > 0$, thereby ensuring that all of the new, shifted, eigenvalues of $M + \alpha Id_n$ are all greater than zero.

2. **(3 points). Using PCA, we reduce the dimension of a dataset $\vec{\mathbf{a}_1}, ..., \vec{\mathbf{a}_n} \in \mathrm{I\!R}^{\mathrm{d}}$ of mean zero, to get a dimensionally reduced dataset $\vec{\mathbf{b}_1}, ..., \vec{\mathbf{b}_n} \in \mathrm{I\!R}^{\mathrm{k}}$, for some $1 \leq k \leq d$.**

   (a) **Show that the dataset $\vec{\mathbf{b}_1}, ..., \vec{\mathbf{b}_n}$ is centered: $\sum_{i=1}^{n} \vec{\mathbf{b}_i} = \vec{\mathbf{0}}$**

   We can express the dimensionally reduced data set $\vec{\mathbf{b}_1}, ..., \vec{\mathbf{b}_n} \in \mathrm{I\!R}^{\mathrm{k}}$ as inner products of our original set with vectors, $\vec{\mathbf{a}_1}, ..., \vec{\mathbf{a}_n} \in \mathrm{I\!R}^{\mathrm{d}}$, with the directions of maximal variance, $\vec{\mathbf{v}_1}, ..., \vec{\mathbf{v}_k} \in \mathrm{I\!R}^{\mathrm{d}}$. In other words, the vectors $\vec{\mathbf{b}_1}, ..., \vec{\mathbf{b}_n} \in \mathrm{I\!R}^{\mathrm{k}}$ can be expressed as,

$$\vec{\mathbf{b}_1}, ..., \vec{\mathbf{b}_n} = \begin{bmatrix} \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_1} \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_1} \rangle \end{bmatrix}, ..., \begin{bmatrix} \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_n} \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_n} \rangle \end{bmatrix} \tag{10}$$

   Then, it becomes clear that,

$$\sum_{i=1}^{n} \vec{\mathbf{b}_i} = \begin{bmatrix} \sum_{i=1}^{n} \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_i} \rangle \\ \vdots \\ \sum_{i=1}^{n} \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_i} \rangle \end{bmatrix} = \begin{bmatrix} \langle \vec{\mathbf{v}_1}, \sum_{i=1}^{n} \vec{\mathbf{a}_i} \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, \sum_{i=1}^{n} \vec{\mathbf{a}_i} \rangle \end{bmatrix} = \begin{bmatrix} \langle \vec{\mathbf{v}_1}, 0 \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, 0 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{\mathbf{0}} \tag{11}$$

   (b) **Show that for all $i, j \in \{1, ..., n\}$, we have**

$$||\vec{\mathbf{b}_i} - \vec{\mathbf{b}_j}|| \leq ||\vec{\mathbf{a}_i} - \vec{\mathbf{a}_j}||$$

   This means that PCA shrinks the distances.

   We can express the difference $\vec{\mathbf{b}_i} - \vec{\mathbf{b}_j}$ as,

$$\vec{\mathbf{b}_i} - \vec{\mathbf{b}_j} = \begin{bmatrix} \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_i} \rangle - \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_j} \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_i} \rangle - \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_j} \rangle \end{bmatrix} = \begin{bmatrix} \langle \vec{\mathbf{v}_1}, \vec{\mathbf{a}_i} - \vec{\mathbf{a}_j} \rangle \\ \vdots \\ \langle \vec{\mathbf{v}_k}, \vec{\mathbf{a}_i} - \vec{\mathbf{a}_j} \rangle \end{bmatrix} \tag{12}$$

   And so it follows that,

$$||\vec{\mathbf{b}_i} - \vec{\mathbf{b}_j}|| = \sqrt{(\vec{\mathbf{v}_1}^T(\vec{\mathbf{a}_i} - \vec{\mathbf{a}_j}))^2 + ... + (\vec{\mathbf{v}_k}^T(\vec{\mathbf{a}_i} - \vec{\mathbf{a}_j}))^2} \tag{13}$$

Similarly, we can express the magnitude of the difference $\overrightarrow{\mathbf{a}_i} - \overrightarrow{\mathbf{a}_j}$ as,

$$||\overrightarrow{\mathbf{a}_i} - \overrightarrow{\mathbf{a}_j}|| = \sqrt{(\overrightarrow{\mathbf{a}_{i_1}} - \overrightarrow{\mathbf{a}_{j_1}})^2 + ... + (\overrightarrow{\mathbf{a}_{i_d}} - \overrightarrow{\mathbf{a}_{j_d}})^2} \tag{14}$$

However, we know that

$$\sum_{e=1}^{k}(\overrightarrow{\mathbf{v}_e}^T(\overrightarrow{\mathbf{a}_i} - \overrightarrow{\mathbf{a}_j}))^2 = (\overrightarrow{\mathbf{a}_{i_1}} - \overrightarrow{\mathbf{a}_{j_1}})^2 + ... + (\overrightarrow{\mathbf{a}_{i_k}} - \overrightarrow{\mathbf{a}_{j_k}})^2 \tag{15}$$

And the final expression becomes,

$$\begin{aligned} ||\overrightarrow{\mathbf{b}_i} - \overrightarrow{\mathbf{b}_j}|| &= \sqrt{(\overrightarrow{\mathbf{a}_{i_1}} - \overrightarrow{\mathbf{a}_{j_1}})^2 + ... + (\overrightarrow{\mathbf{a}_{i_k}} - \overrightarrow{\mathbf{a}_{j_k}})^2} \\ &\leq \sqrt{(\overrightarrow{\mathbf{a}_{i_1}} - \overrightarrow{\mathbf{a}_{j_1}})^2 + ... + (\overrightarrow{\mathbf{a}_{i_d}} - \overrightarrow{\mathbf{a}_{j_d}})^2} = ||\overrightarrow{\mathbf{a}_i} - \overrightarrow{\mathbf{a}_j}|| \end{aligned} \tag{16}$$

Which must be the case since $1 \leq k \leq d$. So we have shown the expression.

(c) For $i \in \{1, ..., k\}$ we let

$$f^{(i)} = (b_{1,i}, b_{2,i}, ..., b_{n,i}) \in \mathrm{I\!R}^{\mathrm{n}}$$

be the vector made of all $i-$th components of the vectors $b_1, ..., b_n$. Show that for $i \neq j$, $f^{(i)} \perp f^{(j)}$. This means that the new features computed using PCA are uncorrelated.

PCA implies that we can represent the covariance matrix, $X \in \mathrm{I\!R}^{\mathrm{k \times k}}$, of some data matrix, call it $A \in \mathrm{I\!R}^{\mathrm{n \times k}}$, as an eigenvalue decomposition such that $X = U \Lambda U^T$, where U contains the eigenvectors corresponding to our new space associated with our principle components. As such, we can represent the data in our new space as $Y = AU$, where every data point is now represented in the basis of the eigenvectors in U.

Alternatively, we can examine the coordinate of every data point in each principle direction. In other words,

$$f^{(i)} = A\overrightarrow{\mathbf{u}_i} \tag{17}$$

Furthermore, it becomes clear that,

$$A\overrightarrow{\mathbf{u}_i} \perp A\overrightarrow{\mathbf{u}_j} \implies f^{(i)} \perp f^{(j)} \tag{18}$$

And, since we know U is an orthonormal matrix, $\overrightarrow{\mathbf{u}_i} \perp \overrightarrow{\mathbf{u}_j}$ for $i \neq j$. So, for $i \neq j$,

$$\langle A\overrightarrow{\mathbf{u}_i}, A\overrightarrow{\mathbf{u}_j} \rangle = \overrightarrow{\mathbf{u}_i}^T A^T A\overrightarrow{\mathbf{u}_j} = \overrightarrow{\mathbf{u}_i}^T U \Lambda U^T \overrightarrow{\mathbf{u}_j} = 0 \tag{19}$$

Which implies that $Au_i \perp Au_j$, and furthermore, that $f^{(i)} \perp f^{(j)}$ for all $i \neq j$.

3. (2 points). Let $A \in \mathbb{R}^{n \times m}$. The Singular Values Decomposition (SVD) tells us that there exists two orthogonal matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ and a matrix $\Sigma \in \mathbb{R}^{n \times m}$ such that $\Sigma_{1,1} \geq \Sigma_{2,2} \geq ... \geq 0$ and $\Sigma_{i,j} = 0$ for $i \neq j$

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

The columns $u_1, ..., u_n$ of U (respectively the columns $v_1, ..., v_m$ of V) are called the left (resp. right) singular vectors of A. The non-negative numbers $\sigma_i = \Sigma_{i,i}$ are the singular values of A. Moreover we also know that $r = rank(A) = num\{i|\Sigma_{i,i} \neq 0\}$.

(a) Let $\tilde{U} = \begin{bmatrix} \vec{\mathbf{u}}_1 & \Big| & ... & \Big| & \vec{\mathbf{u}}_r \end{bmatrix} \in \mathbb{R}^{n \times r}$, $\tilde{V} = \begin{bmatrix} \vec{\mathbf{v}}_1 & \Big| & ... & \Big| & \vec{\mathbf{v}}_r \end{bmatrix} \in \mathbb{R}^{m \times r}$ and $\tilde{\Sigma} = Diag(\sigma_1, ..., \sigma_r) \in \mathbb{R}^{r \times r}$. Show that $A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$

We begin by expressing A in an alternative way - as a sum of rank one matrices. Observe that,

$$\Sigma = D_{1,1} + ... + D_{min(n,m),min(n,m)} \tag{20}$$

Where $D_{i,i}$ is an $n \times m$ matrix where every entry is zero except for the value at index (i, i), where the entry is equal to $\Sigma_{i,i} = \sigma_i$, since we know that $\Sigma_{i,j}$ has a value only when $i = j$, and is 0 otherwise. Therefore, we can express A as,

$$\begin{aligned} A &= U\Sigma V^T = U(D_{1,1} + ... + D_{min(n,m),min(n,m)})V^T \\ &= U(D_{1,1})V^T + ... + U(D_{min(n,m),min(n,m)})V^T \\ &= \sum_{i=1}^{min(n,m)} UD_{i,i}V^T = \sum_{i=1}^{min(n,m)} \sigma_i \vec{\mathbf{u}}_i \vec{\mathbf{v}}_i^T \end{aligned} \tag{21}$$

However, we know that $r = rank(A) = num\{i|\Sigma_{i,i} \neq 0\}$ and that $\Sigma_{1,1} \geq \Sigma_{2,2} \geq ... \geq 0$, meaning any values of $\Sigma_{i,i} = 0$ occur in the final positions of the diagonal, we know that $\Sigma_{r+1,r+1} = ... = \Sigma_{min(n,m),min(n,m)} = 0$. This implies that $\sigma_{r+1} = ... = \sigma_{min(n,m)} = 0$. We then can rewrite the summation as,

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i \vec{\mathbf{u}}_i \vec{\mathbf{v}}_i^T \tag{22}$$

Recomposing this result into matrix form obviously yields $A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$.

(b) Give orthonormal bases of Ker(A) and Im(A) in terms of the singular vectors $u_1, ..., u_n$**,** $v_1, ..., v_m$**.**

The vectors $\overrightarrow{\mathbf{u}_1}, ..., \overrightarrow{\mathbf{u}_r}$ will form a basis for the Im(A). It is evident that the vectors $\overrightarrow{\mathbf{u}_1}, ..., \overrightarrow{\mathbf{u}_r}$ are linearly independent, since U is orthonormal, and furthermore are in the Im(A) since $A\overrightarrow{\mathbf{x}_i} = \overrightarrow{\mathbf{u}_i}$ when $\overrightarrow{\mathbf{x}_i} = \frac{\overrightarrow{\mathbf{v}_i}}{\sigma_i}$ since we have,

$$\overrightarrow{\mathbf{u}_i} = \frac{A\overrightarrow{\mathbf{v}_i}}{\sigma_i} \tag{23}$$

Furthermore, there can only be r vectors in the basis of the Im(A) since $dim(Im(A)) = r$. So the vectors $u_1, ..., u_r$ will form a basis for the Im(A).

By the rank-nullity theorem, the dimension of the kernel is $m - r$. So the vectors $\overrightarrow{\mathbf{v}_{r+1}}, ..., \overrightarrow{\mathbf{v}_m}$ form a basis of the kernel. We know this to be the case because $\overrightarrow{\mathbf{v}_{r+1}}, ..., \overrightarrow{\mathbf{v}_m}$ are linearly independent, since V is orthonormal, and further more are in the Ker(A), since $\overrightarrow{\mathbf{v}_i} = \overrightarrow{\mathbf{0}}$ when $i \in \{r+1, ..., m\}$. This is clearly the case, since,

$$A\overrightarrow{\mathbf{v}_i} = U\Sigma V^T \overrightarrow{\mathbf{v}_i} = \overrightarrow{\mathbf{0}} \tag{24}$$

When $i \in \{r+1, ..., m\}$. Furthermore, there can only be $m - r$ vectors in the basis of the Ker(A) since $dim(Ker(A)) = m - r$. So the vectors $\overrightarrow{\mathbf{v}_{r+1}}, ..., \overrightarrow{\mathbf{v}_m}$ will form a basis for the Ker(A).

4. **(3 points). You have been given a mysterious dataset that may contain important informations! This dataset is a collection of n = 6344 points of dimension d = 1000. Investigate the structure of this dataset using PCA/plots... , and find out if the dataset contains any information.**

The attached Python file fully describes this problem.

In [77]:
```
import numpy as np
D = np.loadtxt(r'mysterious_data.txt')
```

In [78]:
```
## We need to center each column, because each column represents a feature (dimension)

centered_D = D
for i in range(len(D[0,:])):
    centered_D[:,i] = D[:,i] - np.mean(D[:,i])
```

In [79]:
```
## Now we compute the covariance matrix using our newly centered data

cov = np.matmul(centered_D.T,centered_D)
```

In [80]:
```
## Here, we find the eigenvalues and eigenvectors of the covariance matrix

vals, vect = np.linalg.eigh(cov)
```

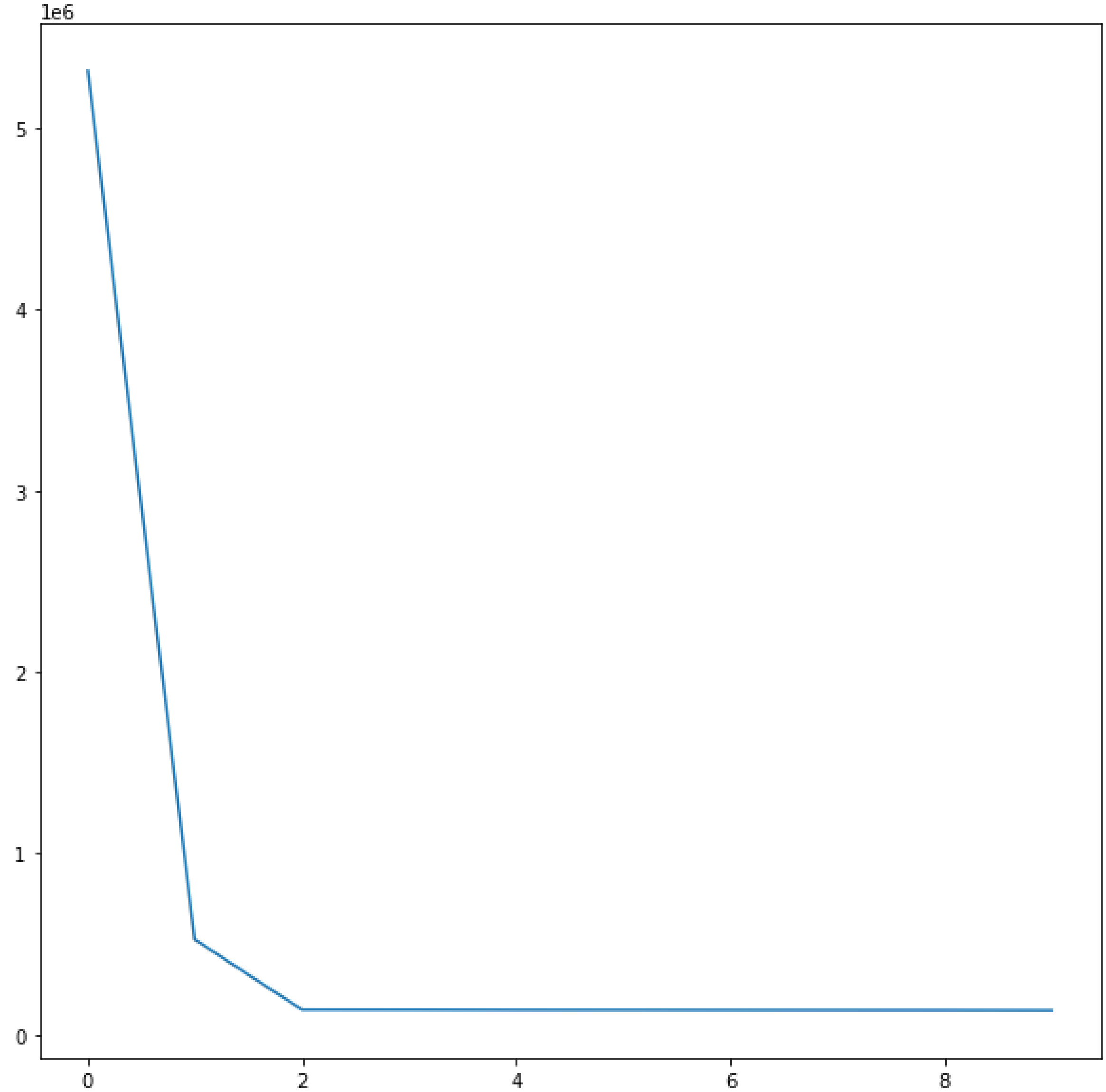
In [81]:
```
## Here we construct the diagonal matrix holding the eigenvalues, and the orthonormal m

E = np.diag(np.flip(vals))
U = vect[:,::-1]
```

In [82]:
```
## This confirms that C = UEU.T

temp = np.matmul(U, E)
check = np.matmul(temp, U.T)
print(check[0][0:5])
print(cov[0][0:5])
```

```
[69258.88218789  1175.26858134   244.45970366  2150.59726671
  2692.39014817]
[69258.88218789  1175.26858134   244.45970366  2150.59726671
  2692.39014817]
```

In [83]:
```
## Here we plot the first 10 eigenvalues to determine significance (only the firs two a

import matplotlib.pyplot as plt
plt.figure(figsize=(10,10))
plt.plot(list(range(len(vals)))[0:10],np.flip(vals)[0:10])
```

Out[83]:
```
[<matplotlib.lines.Line2D at 0x26789349080>]
```

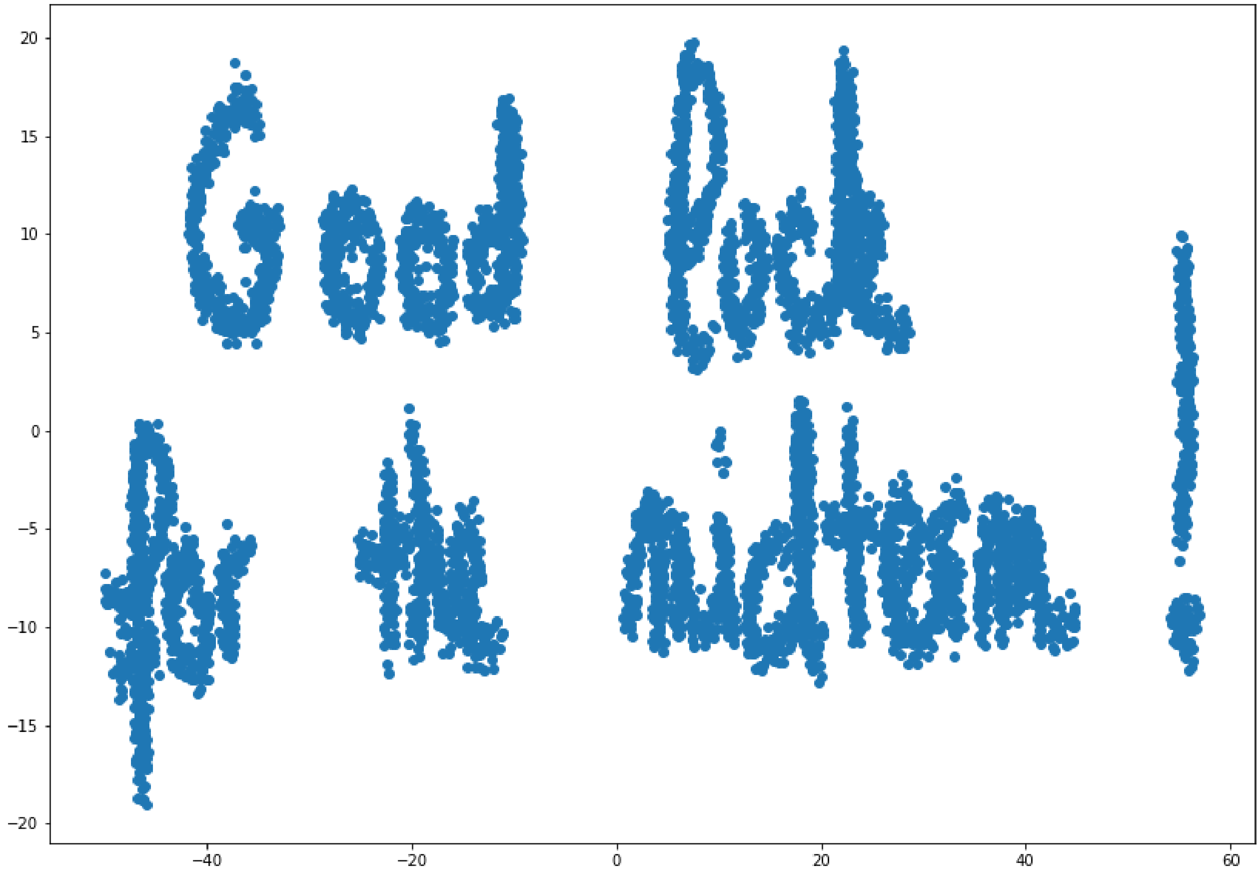


In [84]:
```
## Here, we transform the data into our new space, defined by the orthonormal basis, U

Y = np.matmul(centered_D, U)
```

In [85]:
```
## Here we take the first two dimensions of the data in our new space

ys = [-1*Y[:,0:2][i][0] for i in range(len(Y[:,0:2]))]
xs = [-1*Y[:,0:2][i][1] for i in range(len(Y[:,0:2]))]
```

In [86]:
```
## Here we plot those first two dimensions

plt.figure(figsize=(14,10))
plt.scatter(ys,xs)
```

Out[86]: <matplotlib.collections.PathCollection at 0x267878157f0>


20
15
10
5
0
−5
−10
−15
−20
−40
−20
0
20
40
60