



MSC. DATA SCIENCE & ARTIFICIAL INTELLIGENCE

REFRESHERS: BASIC ALGEBRA FOR DATA ANALYSIS

Dr. Marco CORNELI

---

## Final assignment

---

*Author:* Joris LIMONIER

joris.limonier@hotmail.fr

*Due:* October 22, 2021

# Contents

<b>1</b>	<b>Exercise 1</b>	<b>2</b>
1.1	First part - Question (a) . . . . .	2
1.2	First part - Question (b) . . . . .	3
1.3	Second part - Question (a) . . . . .	3
1.4	Second part - Question (b) . . . . .	4
<b>2</b>	<b>Exercise 2</b>	<b>5</b>
2.1	Question (a) . . . . .	5
2.2	Question (b) . . . . .	6
<b>3</b>	<b>Exercise 3: Principal Component Analysis</b>	<b>6</b>
3.1	First part - Question (a) . . . . .	7
3.2	First part - Question (b) . . . . .	7
3.3	Second part - Question (a) . . . . .	7
3.4	Second part - Question (b) . . . . .	8
3.5	Second part - Question (c) . . . . .	9
3.6	Second part - Question (d) . . . . .	9

# 1 Exercise 1

We consider a symmetric matrix  $S \in \mathbb{R}^{D \times D}$  and the associated quadratic form  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  defined by

$$f(u) := u^T S u = \sum_{i=1}^D \sum_{j=1}^D u_i u_j S_{ij}$$

## 1.1 First part - Question (a)

Let  $i_0 \in \{1, \dots, D\}$ . We want to compute  $\frac{\partial f}{\partial u_{i_0}}(u)$

$$\begin{aligned} \frac{\partial f}{\partial u_{i_0}}(u) &:= \frac{\partial}{\partial u_{i_0}} \left[ \sum_{i=1}^D \sum_{j=1}^D u_i u_j S_{ij} \right] \\ &= \sum_{i=1}^D \sum_{j=1}^D \left[ \frac{\partial}{\partial u_{i_0}} u_i u_j S_{ij} \right] \end{aligned}$$

(Leibniz rule)

$$\begin{aligned} &= \sum_{i=1}^D \sum_{j=1}^D \left[ \frac{\partial u_i}{\partial u_{i_0}} u_j S_{ij} + u_i \frac{\partial u_j}{\partial u_{i_0}} S_{ij} \right] \\ &= \sum_{i=1}^D \sum_{j=1}^D \frac{\partial u_i}{\partial u_{i_0}} u_j S_{ij} + \sum_{i=1}^D \sum_{j=1}^D u_i \frac{\partial u_j}{\partial u_{i_0}} S_{ij} \\ &= \sum_{i=1}^D \underbrace{\frac{\partial u_i}{\partial u_{i_0}}}_{=\mathbf{1}_{\{i=i_0\}}} \sum_{j=1}^D u_j S_{ij} + \sum_{i=1}^D \sum_{j=1}^D u_i \underbrace{\frac{\partial u_j}{\partial u_{i_0}}}_{=\mathbf{1}_{\{j=i_0\}}} S_{ij} \\ &= \sum_{j=1}^D u_j S_{i_0 j} + \sum_{i=1}^D u_i S_{i i_0} \end{aligned}$$

(change counter name  
and since  $S$  is symmetric)

$$\begin{aligned} &= \sum_{i=1}^D u_i S_{i_0 i} + \sum_{i=1}^D u_i S_{i_0 i} \\ &= 2 \sum_{i=1}^D S_{i_0 i} u_i \\ &= 2 S_{i_0 *} u \end{aligned} \tag{1}$$

Where  $S_{i_0*}$  denotes the  $i_0$ -th row of  $S$  (analogously,  $S_{*i_0}$  would denote the  $i_0$ -th column of  $S$ ).

Now by definition, of the gradient of  $f$  and by plugging in the result from (1), we get that:

$$\begin{aligned}\nabla f(u) &:= \begin{bmatrix} \frac{\partial f}{\partial u_1}(u) \\ \vdots \\ \frac{\partial f}{\partial u_D}(u) \end{bmatrix} \\ &= \begin{bmatrix} 2S_{1*}u \\ \vdots \\ 2S_{D*}u \end{bmatrix} \\ &= 2Su\end{aligned}$$

which finishes the proof.

## 1.2 First part - Question (b)

Let  $S := I$ , with  $I \in \mathbb{R}^{D \times D}$  the identity matrix. Naturally,  $I$  is symmetric, therefore we can apply subsection 1.1:

$$\begin{aligned}g(u) &:= u^T u \\ \implies g(u) &= u^T S u \\ \implies g(u) &= f(u) \\ \implies \nabla g(u) &= \nabla f(u) \\ \implies \nabla g(u) &= 2S u \\ \implies \nabla g(u) &= 2u\end{aligned}$$

## 1.3 Second part - Question (a)

Now we assume that  $S$  is positive definite and has eigendecomposition  $S = Q\Lambda Q^T$  with  $Q$  orthogonal and  $\Lambda$  that has the eigenvalues  $\lambda_1, \dots, \lambda_D$  of  $S$  on its diagonal, in decreasing order.

We consider the unit sphere  $\mathcal{U} := \{u \in \mathbb{R}^D \mid \|u\|_2 = 1\}$ .

We want to show that  $\forall u \in \mathcal{U}, y := Q^T u \in \mathcal{U}$ .

$$\begin{aligned}
\|Q^T u\|_2 &= \sqrt{(Q^T u)^T (Q^T u)} \\
&= \sqrt{u^T (Q^T)^T Q^T u} \\
&= \sqrt{u^T Q Q^T u} && (Q Q^T = I \text{ since } Q \text{ is orthogonal}) \\
&= \sqrt{u^T u} \\
&= \sqrt{\|u\|^2} && (\text{Since norms are non-negative}) \\
&= \|u\|
\end{aligned}$$

which completes the proof.

## 1.4 Second part - Question (b)

We want to show the following:

$$\max_{u \in \mathcal{U}} (u^T S u) \leq \lambda_1$$

$$\begin{aligned}
\max_{u \in \mathcal{U}} (u^T S u) &= \max_{u \in \mathcal{U}} (u^T Q \Lambda Q^T u) && (\text{by the order of the eigenvalues}) \\
&\leq \max_{u \in \mathcal{U}} (u^T Q (\lambda_1 I) Q^T u) \\
&&& (\text{since } \lambda_1 \in \mathbb{R}) \\
&= \max_{u \in \mathcal{U}} (\lambda_1 u^T Q Q^T u) \\
&= \max_{u \in \mathcal{U}} (\lambda_1 u^T u) \\
&= \max_{u \in \mathcal{U}} (\lambda_1 \|u\|^2) \\
&&& (\text{since } u \in \mathcal{U}) \\
&= \max_{u \in \mathcal{U}} (\lambda_1) \\
&= \lambda_1
\end{aligned}$$

Hence the proof is complete.

## 2 Exercise 2

### 2.1 Question (a)

$$\begin{aligned}
 P_W(v) &= \sum_{j=1}^K \langle v, u_j \rangle u_j \\
 &= \sum_{j=1}^K u_j \langle v, u_j \rangle \\
 &= \sum_{j=1}^K u_j u_j^T v
 \end{aligned}$$

Therefore, the matrix product looks like:

$$UU^T v = \begin{bmatrix} | & & | \\ u_1 & \dots & u_K \\ | & & | \end{bmatrix} \begin{bmatrix} -u_1^T - \\ \vdots \\ -u_K^T - \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix}$$

where for verification purposes, we can follow the evolution of  $v$  as the linear transformations are applied. This gives us the following visual:

$$\begin{aligned}
 UU^T v &= \begin{bmatrix} | & & | \\ u_1 & \dots & u_K \\ | & & | \end{bmatrix} \begin{bmatrix} -u_1^T - \\ \vdots \\ -u_K^T - \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix} \\
 &= \begin{bmatrix} | & & | \\ u_1 & \dots & u_K \\ | & & | \end{bmatrix} \begin{bmatrix} u_1^T v \\ \vdots \\ u_K^T v \end{bmatrix} \quad (\text{where } \forall 1 \leq i \leq K, u_i^T v \in \mathbb{R}) \\
 &= u_1^T v \begin{bmatrix} | \\ u_1 \\ | \end{bmatrix} + \dots + u_K^T v \begin{bmatrix} | \\ u_K \\ | \end{bmatrix}
 \end{aligned}$$

In the last line, we use the “columns times rows” approach rather than the usual “rows times columns” because it directly yields the result.

We now see that this visual representation confirms the formula obtained previously, so the columns of  $U$  are the  $u_i$ ,  $1 \leq i \leq K$ .

## 2.2 Question (b)

Since  $M_{P_W}$  can be written as  $M_{P_W} = UU^T$ , we have:

$$\begin{aligned}(M_{P_W})^T &= (UU^T)^T \\ &= (U^T)^T U^T \\ &= UU^T \\ &= M_{P_W}\end{aligned}$$

which proves the symmetry.

We use a similar argument to show nilpotence:

$$\begin{aligned}(M_{P_W})^2 &= (UU^T)(UU^T) \\ &= U \underbrace{U^T U}_{=I} U^T \\ &= UU^T \\ &= M_{P_W}\end{aligned}$$

*(since  $\{u_1, \dots, u_K\}$  is an orthonormal basis)*

which proves the nilpotence.

## 3 Exercise 3: Principal Component Analysis

Given  $N$  observations,  $x_1, \dots, x_n \in \mathbb{R}^D$ , we consider their empirical mean

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$$

and their empirical variance

$$\bar{S}_x = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T, \quad \in \mathbb{R}^{D \times D}$$

which is symmetric and positive definite.

### 3.1 First part - Question (a)

Using that  $y_i = U^T x_i$ , we have:

$$\begin{aligned}
 \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i \\
 &= \frac{1}{N} \sum_{i=1}^N U^T x_i \\
 &= U^T \underbrace{\frac{1}{N} \sum_{i=1}^N x_i}_{=\bar{x}} \\
 &= U^T \bar{x}
 \end{aligned}$$

as expected.

### 3.2 First part - Question (b)

$$\begin{aligned}
 U^T \bar{S}_x U &= U^T \left[ \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T \right] U \\
 &= \frac{1}{N} \sum_{i=1}^N \left[ U^T (x_i - \bar{x})(x_i - \bar{x})^T U \right] \\
 &= \frac{1}{N} \sum_{i=1}^N \left[ \left( \underbrace{U^T x_i}_{=y_i} - \underbrace{U^T \bar{x}}_{=\bar{y}} \right) \left( \underbrace{x_i^T U}_{=y_i^T} - \underbrace{\bar{x}^T U}_{=\bar{y}^T} \right) \right] \\
 &= \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})(y_i^T - \bar{y}) \\
 &= \bar{S}_y
 \end{aligned}$$

### 3.3 Second part - Question (a)

We want to show that the maximization of  $\text{trace}(\bar{S}_y)$  can be stated as

$$\max_{\substack{u_1, \dots, u_K \\ u_i \perp u_j}} \sum_{k=1}^K \left[ u_k^T S_x u_k - \lambda_k (\|u_k\|_2^2 - 1) \right], \quad \lambda_k \in \mathbb{R}$$



Let us first compute  $\text{trace}(\bar{S}_y)$ :

$$\begin{aligned}\text{trace}(\bar{S}_y) &= \text{trace}[U^T \bar{S}_x U] \\ &= \sum_{k=1}^K u_k^T \bar{S}_x u_k\end{aligned}\tag{2}$$

but we have to take into account that the  $U$  is an orthogonal matrix, which means:

$$\forall 1 \leq k, l \leq K, u_k^T u_l = \mathbf{1}_{\{k=l\}}$$

hence

$$\forall 1 \leq k \leq K, \begin{cases} \|u_k\|_2^2 = 1 & k = l \\ u_k \perp u_l & k \neq l \end{cases}\tag{3}$$

which implies:

$$\forall 1 \leq k \leq K, \|u_k\|_2^2 - 1 = 0\tag{4}$$

Thus, by equations (2), (3) and (4), the Lagrangian is given by:

$$\mathcal{L}(u_1, \dots, u_K; \lambda_1, \dots, \lambda_K) = \sum_{k=1}^K [u_k^T \bar{S}_x u_k - \lambda_k (\|u_k\|_2^2 - 1)]$$

Finally, maximizing  $\mathcal{L}$  over the  $u_k$ ,  $1 \leq k \leq K$ , which are orthogonal, yields the desired result.

### 3.4 Second part - Question (b)

Assume  $K = 1$  and let  $g(u_1, \lambda_1) := u_1^T \bar{S}_x u_1 - \lambda_1 (\|u_1\|_2^2 - 1)$ . We note that  $g$  is differentiable with respect to all variables since it is a sum of differentiable functions. Thus we get:

$$\nabla g(u_1, \lambda_1) = \begin{bmatrix} 2\bar{S}_x u_1 - 2\lambda_1 u_1 \\ -\|u_1\|_2^2 + 1 \end{bmatrix}$$

which we set equal to 0, therefore obtaining the following system of equations:

$$\begin{aligned}\nabla g(u_1, \lambda_1) &= 0 \\ \implies \begin{cases} 2\bar{S}_x u_1 - 2\lambda_1 u_1 = 0 \\ -\|u_1\|_2^2 + 1 = 0 \end{cases} \\ \implies \begin{cases} \bar{S}_x u_1 = \lambda_1 u_1 \\ \|u_1\|_2^2 = 1 \end{cases}\end{aligned}$$

By definition, we see that the eigenpairs of  $\bar{S}_x$  (with normalized eigenvectors) are solutions of this system, so they are the stationary points of  $g$ .

### 3.5 Second part - Question (c)

The stationary point maximizing  $g$  is the eigenpair (with normalized eigenvector) associated to the largest of the eigenvalues of  $\bar{S}_x$ .

### 3.6 Second part - Question (d)

In the general case, the Lagrangian is maximized by maximizing each of the individual terms of the sum. We can't choose  $K$  repetitions of the eigenpair associated with the largest eigenvalue because of the orthogonality condition, so we choose the eigenpairs associated with the largest  $K$  eigenvalues among the total of all  $D$  eigenvectors of  $\bar{S}_x$ .