Exercise II.1

Let us consider the Shannon entropy function defined as:

$$f(x) = \begin{cases} \sum_{i=1}^{N} x^{(i)} \ln x^{(i)}, & \text{if } (x^{(i)})_{1\le i\le N} \in]0,+\infty[^N \\ +\infty, & \text{if } (\exists j \in \{1,\ldots,N\}) \;\; x^{(j)} < 0. \end{cases} \tag{1}$$

1. How can we extend the definition of f so as to ensure that it is lower semicontinuous on $\mathbb{R}^N$?

2. What can be said about the existence of a minimizer of this function on a nonempty closed subset of the set:

$$C = \left\{ (x^{(i)})_{1\le i\le N} \in [0,+\infty[^N | \sum_{i=1}^{N} x^{(i)} = 1 \right\} \tag{2}$$

Solution: We start by writing $f(x)$ as:

$$f(x) = \sum_{i=1}^{N} \phi\left(x^{(i)}\right), \tag{3}$$

with:

$$\phi(z) = \begin{cases} z \ln z, & \text{if } z \in]0,+\infty[^N \\ +\infty, & \text{if } z < 0. \end{cases}. \tag{4}$$

As a result, since the finite sum of lower semicontinous functions is lower semicontinuous, it suffices to find the value of $\phi(z)$ at $z = 0$ such as to ensure that the function $\phi(z)$ is lower semicontinuous. In order for $\phi(z)$ to be lower semicontinuous, it has to have a closed epigraph. Let us assume that $\phi(0) > 0$ In this case, the epigraph of $\phi(z)$ does not contain all its end points, and as a result it cannot be closed. However, if $\phi(0) \le 0$, the epigraph of $\phi(z)$ containts all its boundary points, and therefore is convex. Therefore, if we extend $\phi(z)$ as:

$$\phi(z) = \begin{cases} z \ln z, & \text{if } z \in]0,+\infty[^N \\ +\infty, & \text{if } z < 0 \\ a, & z = 0, \;\; \text{where } a \le 0. \end{cases}. \tag{5}$$

the function

$$f(x) = \sum_{i=1}^{N} \phi\left(x^{(i)}\right), \tag{6}$$

is lower semicontinuous.

Moving now to the second part of the exercise we note that the set C is closed. Moreover, we also have that C, since it is a subset of $[0,1]^N$, is bounded. As a result, it is also compact. Therefore, using the Weierstrass theorem we have that it has a minimizer.

Exercise II.2

1. Show that the space of real-valued matrices of size $N \times N$ endowed with:

$$\begin{aligned} \langle \cdot | \cdot \rangle : \left(\mathbb{R}^{N \times N}\right)^2 &\mapsto \mathbb{R} \\ (A, B) &\mapsto \operatorname{tr}\left(AB^T\right) \end{aligned} \tag{7}$$

   is a Hilbert space.

2. Let $f : \mathbb{R}^{N \times N} \to \mathbb{R}$ be a continuous function. Does there exist a minimizer of f on the set of orthogonal matrices?

Solution:

1. We start by noticing that the set $\mathbb{R}^{N \times N}$ satisfies all the necessary conditions for being characterized a space. We also investigate the given definition for the given candidate inner product function. After expressing elements of matrix AB^T as:

$$AB^T = \left(\sum_{j=1}^{n} A_{i,j B_{k,j}} \right)_{1 \leq i,k \leq N}, \tag{8}$$

   we can then easily see that:

$$\langle A | B \rangle = \operatorname{tr}\left(AB^T\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{i,j} B_{i,j}. \tag{9}$$

   As a result, if we define the transformation:

$$\begin{aligned} &L : \mathbb{R}^{N \times N} \mapsto \mathbb{R}^{N^2}, \\ &A = (A_{i,j})_{1 \leq i,j \leq N} \mapsto a = (A_{1,1}, \ldots, A_{1,N}, \ldots, A_{N,1}, \ldots, A_{N,N}), \end{aligned} \tag{10}$$

   we can see that L is an isomorphisme between $\mathbb{R}^{N \times N}$ and $\mathbb{R}^{N^2}$ and that the given function is equal to the standard inner product of the euclidean space.

2. Let us consider the set $\mathcal{O}$ of all real $N \times N$ orthogonal matrices. If we prove that this set is a compact set then we can apply the Weierstrass theorem in order to investigate whether or not a function has a minimizer over this set. To do so, let us start by defining a sequence $(A_n)_{n \in \mathbb{N}}$ having a limit point $\bar{A}$. Since for every member of this sequence we have that $A_n A_n^T = I$, it will also hold that $\bar{A}\bar{A}^T = I$. As a result $\bar{A}$ will also be an orthogonal matrix. Hence, we obtain that the limit of a convergent sequence defined in $\mathcal{O}$, converges to a point that also belongs to $\mathcal{O}$. As a result, $\mathcal{O}$ is closed.

Moreover, given the above definition of inner product, we can obtain that $\forall A \in \mathcal{O}$, we have that:

$$\|A\|^2 = \langle A|A\rangle = \mathrm{tr}(AA^T) = \mathrm{tr}(I) = N. \tag{11}$$

As a result the set A is also bounded.

We therefore obtained that $\mathcal{O}$ is finite dmensional, closed and bounded, and therefore compact. Given now that f is continuous, we can apply the Weierstrass theorem and obtain that the function has a minimizer over the set of orthogonal matrices.

Exercise II.3

Let:

$$f : \mathbb{R}^2 \mapsto \mathbb{R}, \quad (u,v) \mapsto \begin{cases} \frac{u^4 v}{u^6+|v|^3}, & \text{if } (u,v) \neq (0,0) \\ 0, & \text{if } (u,v) = 0. \end{cases} \tag{12}$$

1. Is f Gâteaux differentiable at $(0,0)$?
2. Is f continuous at $(0,0)$?
3. Is f Fréchet differentiable at $(0,0)$?

Solution:

1. To investigate whether or not the given function is Gâteaux differentiable, we start by investigating the ratio:

$$\frac{f\left(0+\alpha y\right)-f(0)}{\alpha} = \frac{f\left(\alpha y\right)}{\alpha} \tag{13}$$

We now investigate the following two cases:

- $y = 0$: In this case, based on the definition of f, we can easily obtain that:

$$\frac{f\left(\alpha y\right)}{\alpha} = \frac{f(0)}{\alpha} = 0. \tag{14}$$

- $y = (u,v) \neq 0$: In this case, we have that:

$$\frac{f\left(\alpha y\right)}{\alpha} == \frac{\alpha^5 u^4 v}{\alpha\left(\alpha^6 u^6 + \alpha^3 |v|^3\right)} = \alpha\left(\frac{u^4}{a^3 u^6 + |v|^3}\right) \to 0, \text{ for } \alpha \to 0. \tag{15}$$

As a result, we obtain that $\forall y$ as $\alpha \to 0$, the ratio $\frac{f(\alpha y)}{\alpha} = \frac{f(0)}{\alpha} = 0.$ converges to the same value, i.e. to the value of zero. As a result, the function is Gâteaux differentiable at zero.

2. We now investigate whether or not the function is continuous at zeor. To do so, we start by investigating the value of $f(y)$ as $y \to 0$. In particular, we consider vectors of the form $y = (u, u^2), u \neq 0$. For this set of vectors, we can obtain that:

$$f\left(u, u^2\right) = \frac{u^6}{2u^6} = 0.5. \tag{16}$$

   However, this value is different from $f(0,0)$. As a result we cannot guarantee that $f(y)$ converges to a value as y converges to zero. Therefore, the function is not continuous.

3. Since f is not continuous at $(0,0)$, it follows that is also not Fréchet differentiable at $(0,0)$.

Exercise II.4

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ and $f : \mathcal{G} \to \mathbb{R}$.

1. If $x \in \mathcal{H}$ and f is Gâteux differentiable at Lx, show that $g = f \odot L$ is Gâteux differentiable at x and that $\nabla g(x) = L^* \nabla f(Lx)$.

2. If $x \in \mathcal{H}$ and f us Fréchet differentiable at Lxn do we have the same conclusion?

3. If $x \in \mathcal{H}$ and f is twice Fréchet differentiable at Lx, show that g is twice Fréchet differentiable at x and

$$\nabla^2 g\left(x\right) = L^* \nabla^2 f(Lx) L \tag{17}$$

Solution

1. In order to investigate if g is Gâteaux differentiable, at xn we need to find, $\forall y \in \mathcal{H}$ the limit:

$$\lim_{\alpha \to 0} \frac{g\left(x + ay\right) - g(x)}{\alpha}. \tag{18}$$

   Using the connection between f and g, the above limit is expressed as:

$$\begin{aligned} \lim_{\alpha \to 0} \frac{g\left(x + ay\right) - g(x)}{\alpha} &= \lim_{\alpha \to 0} \frac{f\left(Lx + aLy\right) - f(Lx)}{\alpha} \\ &= \langle \nabla f(Lx), Ly \rangle = \langle L^* \nabla f(Lx), y \rangle. \end{aligned} \tag{19}$$

2. In order to investigate Fréchet differentiability, we need to consider the limit:

$$\lim_{y \to 0, y \neq 0} \frac{g\left(x + y\right) - g(x) - \langle \nabla f(Lx), Ly \rangle}{\|y\|} = \lim_{y \to 0, y \neq 0} \frac{g\left(x + y\right) - g(x) - \langle L^* \nabla f(Lx), y \rangle}{\|y\|} \tag{20}$$

   Let us now consider separately the following two cases:

- $Ly = 0$. In this case we obtain that $f(L(x + y)) = f(Lx)$ and $\langle \nabla f(Lx), Ly \rangle = 0$, and we reach the conclusion that

$$\lim_{y \to 0, y \neq 0} \frac{g\,(x + y) - g(x) - \langle \nabla f(Lx), Ly \rangle}{\|y\|} = 0 \tag{21}$$

- $Ly \neq 0$: In this case we can write:

$$\frac{|g\,(x + y) - g(x) - \langle \nabla f(Lx), Ly \rangle|}{\|y\|} = \frac{\|Ly\|}{\|y\|} \frac{|f\,(Lx + Ly) - f(Lx) - \langle \nabla f(Lx), Ly \rangle|}{\|y\|} \tag{22}$$

Focusing now on this product representation, due to the fact that $L \in \mathcal{B}\,(\mathcal{H}, \mathcal{G})$ we have that the first term is bounded. Moreover, we also have that if f is Fréchet differentiable, as $y \to 0$, the second term will converge to zero. We therefore obtain that:

$$\lim_{y \to 0, y \neq 0} \frac{g\,(x + y) - g(x) - \langle \nabla f(Lx), Ly \rangle}{\|y\|} = 0, \tag{23}$$

which means that g is also Fréchet differentiable.

3. Essentially, we need to prove that:

$$\lim_{y \to 0, y \neq 0} \frac{\nabla g(x + y) - \nabla g(x) - L^* \nabla^2 f(Lx) L}{\|y\|} = 0. \tag{24}$$

To do so, we start by using the expresson for ∇g derived earlier and rewrite the given limit as:

$$\begin{aligned} &\lim_{y \to 0, y \neq 0} \frac{\nabla g(x + y) - \nabla g(x) - L^* \nabla^2 f(Lx) L}{\|y\|} \\ &= \lim_{y \to 0, y \neq 0} \frac{L^* \nabla f(L(x + y)) - L^* \nabla f(Lx) - L^* \nabla^2 f(Lx) L}{\|y\|} \\ &= \lim_{y \to 0, y \neq 0} L^* \frac{\nabla f(L(x + y)) - \nabla f(Lx) - \nabla^2 f(Lx) L}{\|y\|} \end{aligned} \tag{25}$$

Using the same process as above, we can then prove that this limit is equal to zero. Moreover, we transform $L^* \nabla^2 f(Lx) Ly$ is bounded, linear and continuous. As a result, the function g is twice Fréchet differentiable, and the value of the derivative is equal to the given one.