Exercise 11

Due date:

Name : Vivi
Student ID : 24S153073
Grade : ________

Problem 1 Score: ________. Properties of g-convex sets and functions
Let $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold and consider a geodesically convex subset $S \subseteq \mathcal{M}$. Let $f, f_1, \cdots, f_n : S \to \mathbb{R}$ be g-convex functions.

(1) Sublevel sets of g-convex functions are g-convex sets.
Let $s \in \mathbb{R}$. Show that the sublevel set
$$L_f(s) = \{x \in S : f(x) \leq s\}$$
is g-convex.

(2) Intersections of sublevel sets are g-convex sets.
Let $s_1, s_2 \in \mathbb{R}$. Show that the intersection
$$\bigcap_{j=1}^{n} L_{f_j}(s_j)$$
is g-convex.

(3) Sums of nonnegatively scaled g-convex functions are g-convex functions.
Let $\alpha_1, \cdots, \alpha_n \geq 0$. Show that the function
$$g(x) = \sum_{j=1}^{n} \alpha_j f_j(x)$$
is g-convex.

(4) The pointwise maximum of g-convex functions is g-convex.
Show that the function
$$h(x) = \max_{j=1,\cdots,n} f_j(x)$$
is g-convex.

Solution: (1) Let $x, y \in L_f(s)$ and $\gamma : [0,1] \to S$ be a geodesic segment with $\gamma(0) = x$ and $\gamma(1) = y$. Since S is g-convex, we have $\gamma(t) \in S$ for all $t \in [0,1]$. By the g-convexity of f, for all $t \in [0,1]$, we have
$$\begin{aligned} f \circ \gamma(t) &\leq (1-t)f(x) + tf(y) \\ &\leq (1-t)s + ts = s, \end{aligned}$$
which implies that $\gamma(t) \in L_f(s)$ for all $t \in [0,1]$. Therefore, $L_f(s)$ is g-convex.

(2) Let $x, y \in \bigcap_{j=1}^{n} L_{f_j}(s_j)$ and $\gamma : [0,1] \to S$ be a geodesic segment with $\gamma(0) = x$ and $\gamma(1) = y$. Since S is g-convex, we have $\gamma(t) \in S$ for all $t \in [0,1]$. By the g-convexity of f_j, for all $t \in [0,1]$, we have
$$\begin{aligned} f_j \circ \gamma(t) &\leq (1-t)f_j(x) + tf_j(y) \\ &\leq (1-t)s_j + ts_j = s_j \end{aligned}$$
which implies that $\gamma(t) \in L_{f_j}(s_j)$ for all $j = 1, \cdots, n \quad \forall t \in [0,1]$, then $\gamma(t) \in \bigcap_{j=1}^{n} L_{f_j}(s_j)$. Therefore $\bigcap_{j=1}^{n} L_{f_j}(s_j)$ is g-convex.

(3) Let $x, y \in S$ and $\gamma : [0,1] \to S$ be a geodesic segment with $\gamma(0) = x$ and $\gamma(1) = y$. Since S is g-convex, we have $\gamma(t) \in S$ for all $t \in [0,1]$. By the g-convexity of f_j, for all $t \in [0,1]$, we have
$$\begin{aligned} g \circ \gamma(t) &= \sum_{j=1}^{n} \alpha_j f_j \circ \gamma(t) \\ &\leq \sum_{j=1}^{n} \alpha_j((1-t)f_j(x) + tf_j(y)) \\ &= (1-t)\sum_{j=1}^{n} \alpha_j f_j(x) + t\sum_{j=1}^{n} \alpha_j f_j(y) \\ &= (1-t)g(x) + tg(y) \end{aligned}$$
for all $t \in [0,1]$. Therefore, g is g-convex.

(4) Let $x, y \in S$ and $\gamma : [0,1] \to S$ be a geodesic segment with $\gamma(0) = x$ and $\gamma(1) = y$. Since S is g-convex, we have $\gamma(t) \in S$ for all $t \in [0,1]$. By the definition of h, we have

$$\begin{aligned} h \circ \gamma(t) &= \max_{j=1,\cdots,n} f_j \circ \gamma(t) \\ &\leq \max_{j=1,\cdots,n} ((1-t) f_j(x) + t f_j(y)) \\ &\leq (1-t) \max_{j=1,\cdots,n} f_j(x) + t \max_{j=1,\cdots,n} f_j(y) \\ &= (1-t) h(x) + t h(y) \end{aligned}$$

for all $t \in [0,1]$. Therefore, h is g-convex.

□

Problem 2 Score: ________. Intersection of g-convex sets
The intersection of two convex subsets of a Euclidean space is convex. However, in general, the intersection of two g-convex sets is not g-convex.

(1) Give an example of a Riemannian manifold $(\mathcal{M}, \langle \cdot, \cdot \rangle)$ and two g-convex sets $S_1, S_2 \subseteq \mathcal{M}$ such that $S_1 \cap S_2$ is not g-convex.

If we make additional assumptions, then the intersection of two g-convex sets is g-convex. A subset $S \subseteq \mathcal{M}$ is geodesically strongly convex if for any two points $x, y \in S$, among all geodesics segments $\gamma : [0,1] \to S$ with $\gamma(0) = x$ and $\gamma(1) = y$, exactly one of them is minimizing and this minimizing geodesic lies entirely in S.

(2) Let $S_1, S_2 \subseteq \mathcal{M}$ be two geodesically strongly convex sets. Show that $S_1 \cap S_2$ is geodesically strongly convex.

Solution: (1) Let $\mathcal{M} = \mathbb{S}^1$ be the unit circle in $\mathbb{R}^2$ with the standard metric. Define the following two g-convex sets:

$$\begin{aligned} S_1 &= \mathbb{S}^1 \setminus \{(0,1)\} \\ S_2 &= \mathbb{S}^1 \setminus \{(0,-1)\}. \end{aligned}$$

Then S_1 and S_2 are g-convex sets. However, the intersection $S_1 \cap S_2 = \mathbb{S}^1 \setminus \{(0,1), (0,-1)\}$ is not g-convex because it's not connected.

(2) Let $x, y \in S_1 \cap S_2$ and $\gamma : [0,1] \to S_1 \cap S_2$ be the unique geodesic segment with $\gamma(0) = x$ and $\gamma(1) = y$. Since S_1, S_2 are geodesically strongly convex, $\gamma(t) \in S_1$ and $\gamma(t) \in S_2$ for all $t \in [0,1]$. Therefore, $\gamma(t) \in S_1 \cap S_2$ for all $t \in [0,1]$, which implies that $S_1 \cap S_2$ is geodesically strongly convex.

□

Problem 3 Score: ________. Fréchet mean on hemisphere
Write some code to generate random points $x_1, \cdots, x_n$ on a hemisphere

$$\mathbb{S}_+^{d-1} := \{x = (x^{(1)}, \cdots, x^{(d)}) \in \mathbb{R}^d : x^{(d)} > 0, \|x\| = 1\}$$

near the north pole, and implement the cost function for the intrinsic averaging, that is

$$f : \mathbb{S}_+^{d-1} \to \mathbb{R}, \quad f(x) = \frac{1}{2n} \sum_{i=1}^{n} \operatorname{dist}(x, x_i)^2.$$

A global minimizer of f is called the Fréchet mean of $x_1, \cdots, x_n$.
Recall that the squared distance between two points $x, y \in \mathbb{S}_+^{d-1}$ is given by

$$\operatorname{dist}(x,y)^2 = \arccos^2(x^\top y),$$

and the Riemannian gradient of the squared distance is given by

$$\operatorname{grad}\left(x \mapsto \frac{1}{2}\operatorname{dist}(x,y)\right)(x) = \frac{\operatorname{dist}(x,y)}{\sin(\operatorname{dist}(x,y))}(\cos(\operatorname{dist}(x,y))x - y).$$

Problem 4 Score: ________. Robust covariance estimation
Consider n points $x_1, \cdots, x_n \in \mathbb{R}^d$ sampled independently and identically distributed from a distribution P with zero

mean. We want to estimate the covariance matrix of P. If P is a zero-mean normal distribution with covariance $\Sigma_{true} \in \mathbb{R}^{d\times d}$, then the maximum likelihood estimation amounts to minimizing the negative log-likelihood

$$\Sigma \mapsto \log(\det \Sigma) + \frac{1}{n}\sum_{j=1}^{n} x_j^\top \Sigma^{-1} x_j$$

over the $d \times d$ positive definite matrices

$$\mathcal{P}_d = \{\Sigma \in \mathbb{R}^{d\times d} : \Sigma = \Sigma^\top, \Sigma \succ 0\}.$$

The sample covariance matrix $\Sigma^* = \frac{1}{n}\sum_{j=1}^{n} x_j x_j^\top$ is a minimizer of this nagetive log-likelihood.
The sample covariance is not robust to outliers. So if P is not normal but some heavy-tailed distribution, then the sample covariance is not suitable. We can obtain a robust estimation of the covariance by minimizing the function

$$f : \mathcal{P}_d \to \mathbb{R}, \quad f(\Sigma) = \log(\det \Sigma) + \frac{1}{n}\sum_{j=1}^{n} d\log(x_j^\top \Sigma^{-1} x_j),$$

which places less emphasis on outliers (points far from the mean). A minimizer of this function is called "Tyler's M-estimator of scatter". It does not have a closed form solution, and the cost function f is non-convex in the Euclidean sense. However, it is g-convex in an appropriate metric, and so a minimizer can be found efficiently(e.g., with RGD).
We consider $\mathcal{M} = \mathcal{P}_d$ as an open subset of the symmetric $d \times d$ matrices, and endow it with the Fisher-Rao metric

$$\langle \dot\Sigma_1, \dot\Sigma_2\rangle_\Sigma = \mathrm{Tr}(\Sigma^{-1}\dot\Sigma_1\Sigma^{-1}\dot\Sigma_2),$$

for $\Sigma \in \mathcal{P}_d$ and $\dot\Sigma_1, \dot\Sigma_2 \in \mathrm{T}_\Sigma\mathcal{P}_d = \{\dot\Sigma \in \mathbb{R}^{d\times d} : \dot\Sigma = \dot\Sigma^\top\}$. In this Riemannian metric, $\mathcal{P}_d$ is complete and geodesically strongly convex. For every $\Sigma_0, \Sigma_1 \in \mathcal{P}_d$, there is a unique geodesic segment between them, given by

$$\gamma(t) = \Sigma_0^{1/2}(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})^t\Sigma_0^{1/2}, \quad t \in [0,1].$$

This geodesic segment is minimizing. Alternatively, for every $\Sigma_0, \Sigma_1 \in \mathcal{P}_d$, there exists an invertible $V \in \mathbb{R}^{d\times d}$ and a diagonal $D \in \mathcal{P}_d$ such that $\Sigma_0 = VV^\top, \Sigma_1 = VDV^\top$. In this case,

$$\gamma(t) = VD^tV^\top, \quad t \in [0,1].$$

(1) Show that the function $\Sigma \mapsto \log(\det \Sigma)$ is g-convex.

(2) Show that if $g : \mathcal{P}_d \to \mathbb{R}$ is g-convex, then the function $h(\Sigma) = g(\Sigma^{-1})$ is g-convex.

(3) Show that if $x \in \mathbb{R}^d$, then the function $\Sigma \mapsto \log(x^\top\Sigma x)$ is g-convex.

(4) Conclude that the function f is g-convex.

Solution: (1) Let $\Sigma_0, \Sigma_1 \in \mathcal{P}_d$, then the unique geodesic segment between Σ_0 and Σ_1 is given by

$$\gamma(t) = \Sigma_0^{1/2}(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})^t\Sigma_0^{1/2}, \quad t \in [0,1].$$

For all $t \in [0,1]$, we have

$$\begin{aligned}
\log(\det \gamma(t)) &= \log(\det(\Sigma_0^{1/2}(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})^t\Sigma_0^{1/2})) \\
&= \log\left(\det(\Sigma_0^{1/2})\det(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})^t\det(\Sigma_0^{1/2})\right) \\
&= \log(\det(\Sigma_0)) + t\log(\det(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})) \\
&= \log(\det(\Sigma_0)) + t\log\left(\det(\Sigma_0^{-1/2})\det(\Sigma_1)\det(\Sigma_0^{-1/2})\right) \\
&= \log(\det(\Sigma_0)) + t\log(\det(\Sigma_1)) - t\log(\det(\Sigma_0)) \\
&= (1-t)\log(\det(\Sigma_0)) + t\log(\det(\Sigma_1))
\end{aligned}$$

Therefore, $\log(\det \Sigma)$ is g-convex. Moreover, it's g-affine.

(2) Let $\Sigma_0, \Sigma_1 \in \mathcal{P}_d$, then the unique geodesic segment between Σ_0 and Σ_1 is given by

$$\gamma(t) = \Sigma_0^{1/2}(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})^t\Sigma_0^{1/2}, \quad t \in [0,1].$$

Let $g : \mathcal{P}_d \to \mathbb{R}$ be g-convex, then for all $t \in [0,1]$, we have

$$\begin{aligned} h(\gamma(t)) &= g(\gamma(t)^{-1}) \\ &\leq (1-t)g(\Sigma_0^{-1}) + tg(\Sigma_1^{-1}) \\ &= (1-t)h(\Sigma_0) + th(\Sigma_1). \end{aligned}$$

Therefore, h is g-convex.

(3) Fix $x \in \mathbb{R}^d \setminus \{0\}$ and $F : \mathcal{P}_d \to \mathbb{R}$ given by $F(\Sigma) = \log(x^\top \Sigma x)$. Let $\Sigma \in \mathcal{P}_d$ and $\dot{\Sigma} \in \mathrm{T}_\Sigma\mathcal{P}_d$, then the differential of F at Σ in the direction $\dot{\Sigma}$ is given by

$$\begin{aligned} DF(\Sigma)(\dot{\Sigma}) &= \frac{d}{dt}\log(x^\top(\Sigma + t\dot{\Sigma})x)\bigg|_{t=0} \\ &= (x^\top\Sigma x)^{-1}x^\top\dot{\Sigma}x \\ &= (x^\top\Sigma x)^{-1}\operatorname{Tr}(\dot{\Sigma}xx^\top). \end{aligned}$$

Then the Euclidean gradient of F at Σ is given by

$$\operatorname{grad}_\mathcal{E} F(\Sigma) = \frac{xx^\top}{x^\top\Sigma x}.$$

The Euclidean Hessian of F at Σ is given by

$$\begin{aligned} \operatorname{Hess}_\mathcal{E} F(\Sigma)(\dot{\Sigma}) &= D\operatorname{grad}_\mathcal{E} F(\Sigma)(\dot{\Sigma}) \\ &= D\left(\frac{xx^\top}{x^\top\Sigma x}\right)(\dot{\Sigma}) \\ &= -\frac{xx^\top x^\top\dot{\Sigma}x}{(x^\top\Sigma x)^2}. \end{aligned}$$

Then the Riemannian Hessian of F at Σ is given by

$$\begin{aligned} \operatorname{Hess}_\mathcal{M} F(\Sigma)(\dot{\Sigma}) &= \Sigma\operatorname{Hess}_\mathcal{E} F(\Sigma)(\dot{\Sigma})\Sigma + \frac{\dot{\Sigma}\operatorname{grad}_\mathcal{E} F(\Sigma)\Sigma + \Sigma\operatorname{grad}_\mathcal{E} F(\Sigma)\dot{\Sigma}}{2} \\ &= -\frac{\Sigma xx^\top x^\top\dot{\Sigma}x\Sigma}{(x^\top\Sigma x)^2} + \frac{\dot{\Sigma}xx^\top\Sigma + \Sigma xx^\top\dot{\Sigma}}{2x^\top\Sigma x} \\ &= \frac{1}{2x^\top\Sigma x}(\dot{\Sigma}xx^\top\Sigma + \Sigma xx^\top\dot{\Sigma} - \frac{2\Sigma xx^\top x^\top\dot{\Sigma}x\Sigma}{x^\top\Sigma x}). \end{aligned}$$

To show the Riemannian Hessian is positive semidefinite,

$$\begin{aligned} \langle\operatorname{Hess}_\mathcal{M} F(\Sigma)(\dot{\Sigma}), \dot{\Sigma}\rangle_\Sigma &= \operatorname{Tr}\left(\Sigma^{-1}\operatorname{Hess}_\mathcal{M} F(\Sigma)(\dot{\Sigma})\Sigma^{-1}\dot{\Sigma}\right) \\ &= \frac{1}{2x^\top\Sigma x}\operatorname{Tr}\left(\Sigma^{-1}(\dot{\Sigma}xx^\top\Sigma + \Sigma xx^\top\dot{\Sigma} - \frac{2\Sigma xx^\top x^\top\dot{\Sigma}x\Sigma}{x^\top\Sigma x})\Sigma^{-1}\dot{\Sigma}\right) \\ &= \frac{1}{2x^\top\Sigma x}\operatorname{Tr}\left(\Sigma^{-1}\dot{\Sigma}xx^\top\dot{\Sigma} + xx^\top\dot{\Sigma}\Sigma^{-1}\dot{\Sigma} - \frac{2xx^\top x^\top\dot{\Sigma}x\dot{\Sigma}}{x^\top\Sigma x}\right) \\ &= \frac{1}{x^\top\Sigma x}\left(x^\top\dot{\Sigma}\Sigma^{-1}\dot{\Sigma}x - \frac{(x^\top\dot{\Sigma}x)^2}{x^\top\Sigma x}\right) \\ &= \frac{1}{(x^\top\Sigma x)^2}\left[(x^\top\Sigma x)(x^\top\dot{\Sigma}\Sigma^{-1}\dot{\Sigma}x) - (x^\top\dot{\Sigma}x)^2\right] \geq 0. \end{aligned}$$

Therefore, $F(\Sigma) = \log(x^\top\Sigma x)$ is g-convex.

(4) As shown in (1), (2), and (3), the functions $\log(\det\Sigma)$, $\log(\det\Sigma^{-1})$, and $\log(x^\top\Sigma x)$ are g-convex. Therefore, the function f is g-convex as a non-negative combination of g-convex functions.

□