Exercise 7

Due date:

Name : Vivi
Student ID : 24S153073
Grade : ________

Problem 1 Score: ________. Reimannian Hessian on Stiefel
For $p \leq n$, consider the Stiefel manifold

$$\mathcal{M} = \mathrm{St}(n, p) = \{X \in \mathbb{R}^{n\times p} : X^\top X = I_p\}$$

as an embedded submanifold of $\mathcal{E} = \mathbb{R}^{n\times p}$. We consider $\mathcal{M}$ as a Riemannian manifold of $\mathcal{E} = \mathbb{R}^{n\times p}$, endowed with the usual inner product $\langle X, Y\rangle = \mathrm{Tr}(X^\top Y)$.
The orthogonal projection to $T_X\mathcal{M}$ is given by

$$\mathrm{Proj}_X : \mathcal{E} \to T_X\mathcal{M}, \quad \mathrm{Proj}_X(U) = U - \frac{1}{2}X(X^\top U + U^\top X) = U - X\mathrm{Sym}(X^\top U),$$

where $\mathrm{Sym}(A) = \frac{1}{2}(A + A^\top)$ is the symmetrization of A.
Let $f : \mathcal{M} \to \mathbb{R}$ be smooth, and let $\bar{f}$ be a smooth extension of f.

(1) Give a formula for the Riemannian Hessian $\mathrm{Hess}f$ of f in terms of the Euclidean gradient and Hessian of $\bar{f}$.

Let R be a retraction on Stiefel (e.g., QR or polar retraction). Let $(X, U) \in T\mathcal{M}$, the finite difference approximation of the Riemannian Hessian is given by

$$\mathrm{Hess}f(X)[U] \approx \frac{1}{\bar{t}}[\mathrm{Proj}(\mathrm{grad}\, f(R_X(\bar{t}U))) - \mathrm{grad}\, f(X)]$$

where $\bar{t} > 0$ is a small step size.

(2) For the particular cost function

$$f(X) = \mathrm{Tr}(X^\top AX), \quad A \in \mathbb{R}^{n\times n} \text{with} A = A^\top,$$

write down a formula for the Riemannian Hessian of f, and a formula for the finite difference approximation of the Riemannian Hessian. Implement both formulas, and compare them for different values of $\bar{t}$ (e.g., $\bar{t} = 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}$).

Solution: (1) The Riemannian gradient of f is given by

$$\begin{aligned}\mathrm{grad}\, f(X) &= \mathrm{Proj}_X(\mathrm{grad}\, \bar{f}(X)) \\ &= \mathrm{grad}\, \bar{f}(X) - X\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X)),\end{aligned}$$

which can be smoothly extended as

$$\overline{\mathrm{grad}\, f}(X) = \mathrm{grad}\, \bar{f}(X) - X\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X)).$$

The derivative of $\overline{\mathrm{grad}\, f}$ is given by

$$\begin{aligned}D\overline{\mathrm{grad}\, f}(X)[U] &= D(\mathrm{grad}\, \bar{f}(X) - X\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X)))[U] \\ &= D\,\mathrm{grad}\, \bar{f}(X)[U] - D(X\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X)))[U] \\ &= \mathrm{Hess}\, \bar{f}(X)[U] - U\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X)) - X\mathrm{Sym}(U^\top \mathrm{grad}\, \bar{f}(X) + X^\top \mathrm{Hess}\, \bar{f}(X)[U]).\end{aligned}$$

Observes that

$$\begin{aligned}\mathrm{Proj}_X(XS) &= XS - X\mathrm{Sym}(X^\top XS) \\ &= XS - X\mathrm{Sym}(S) = 0\end{aligned}$$

for any symmetric $S \in \mathbb{R}^{p\times p}$.
Then, the Riemannian Hessian of f is given by

$$\begin{aligned}\mathrm{Hess}\, f(X)[U] &= \nabla_U\, \mathrm{grad}\, f(X) \\ &= \mathrm{Proj}_X(D\overline{\mathrm{grad}\, f}(X)[U]) \\ &= \mathrm{Proj}_X(\mathrm{Hess}\, \bar{f}(X)[U]) - \mathrm{Proj}_X(U\mathrm{Sym}(X^\top \mathrm{grad}\, \bar{f}(X))).\end{aligned}$$

(2) Define $\bar{f}(X) = \mathrm{Tr}(X^\top AX)$, then for $(X, U) \in T\mathcal{M}$, we have

$$\begin{aligned} \mathrm{grad}\, \bar{f}(X) &= 2AX \\ \mathrm{Hess}\, \bar{f}(X)[U] &= 2AU. \end{aligned}$$

Thus, the Riemannian Hessian of f is given by

$$\begin{aligned} \mathrm{Hess}\, f(X)[U] &= \mathrm{Proj}_X(2AU) - \mathrm{Proj}_X(U\mathrm{Sym}(X^\top 2AX)) \\ &= 2\,\mathrm{Proj}_X(AU) - 2\,\mathrm{Proj}_X(UX^\top AX). \end{aligned}$$

The Riemannian gradient of f is given by

$$\begin{aligned} \mathrm{grad}\, f(X) &= \mathrm{Proj}_X(2AX) \\ &= 2AX - X\mathrm{Sym}(X^\top 2AX) \\ &= 2AX - 2XX^\top AX \\ &= 2(I_n - XX^\top)AX. \end{aligned}$$

□

Problem 2 Score: _________. Second-order critical points for Rayleigh quotient are global optimal

Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth Riemannian manifold $\mathcal{M}$. We call $x \in \mathcal{M}$ a second-order critical point of f if

$$\mathrm{grad}\, f(x) = 0 \quad \text{and} \quad \mathrm{Hess}\, f(x) \succeq 0.$$

Let $\mathcal{M} = \mathbb{S}^{d-1}$ be the $(d-1)$-dimensional sphere embedded in $\mathbb{R}^d$ with the usual inner product, and let

$$f(x) = \frac{1}{2}x^\top Ax,$$

where symmetric $A \in \mathbb{R}^{d\times d}$. This cost function is sometimes called the Rayleigh quotient on the sphere.

(1) Give expressions for the Riemannian gradient and Hessian of f.

(2) Show that all second-order critical points x of f are globally optimal.

Solution: (1) We've already known that the projection to $T_x\mathcal{M}$ is given by

$$\mathrm{Proj}_x(u) = (I - xx^\top)u.$$

The Euclidean gradient and Hessian of $\bar{f} = \frac{1}{2}x^\top Ax$ are given by

$$\begin{aligned} \mathrm{grad}\, \bar{f}(x) &= Ax \\ \mathrm{Hess}\, \bar{f}(x)[u] &= Ax. \end{aligned}$$

Thus, the Riemannian gradient of f are given by

$$\begin{aligned} \mathrm{grad}\, f(x) &= \mathrm{Proj}_x(\mathrm{grad}\, \bar{f}(x)) \\ &= (I - xx^\top)Ax, \end{aligned}$$

which can be smoothly extended as

$$\overline{\mathrm{grad}\, f}(x) = (I - xx^\top)Ax.$$

The derivative of $\overline{\mathrm{grad}\, f}$ is given by

$$\begin{aligned} D\overline{\mathrm{grad}\, f}(x)[u] &= D((I - xx^\top)Ax)[u] \\ &= (I - xx^\top)Au - (ux^\top + xu^\top)Ax. \end{aligned}$$

Then, the Riemannian Hessian of f is given by

$$\begin{aligned} \mathrm{Hess}\, f(x)[u] &= \mathrm{Proj}_x(D\overline{\mathrm{grad}\, f}(x)[u]) \\ &= \mathrm{Proj}_x((I - xx^\top)Au - (ux^\top + xu^\top)Ax) \end{aligned}$$

$$
\begin{aligned}
&= (I - xx^\top)Au - (ux^\top + xu^\top)Ax - xx^\top Au + ux^\top Axx^\top Ax \\
&= (I - xx^\top)Au - (ux^\top + xu^\top)Ax \\
&= Au - 2(x^\top Au)x - (x^\top Ax)u.
\end{aligned}
$$

Then, the Riemannian Hessian of f is given by

$$
\begin{aligned}
\operatorname{Hess} f(x)[u] &= \nabla_u \operatorname{grad} f(x) \\
&= \operatorname{Proj}_x(D\overline{\operatorname{grad} f}(x)[u]) \\
&= \operatorname{Proj}_x(Au) - x^\top Ax \\
&= (I - xx^\top)Au - (x^\top Ax)u
\end{aligned}
$$

(2) For critical points x of f, we have $\operatorname{grad} f(x) = (I - xx^\top)Ax = 0$, which implies $Ax = (x^\top Ax)x$, i.e., x is an eigenvector of A with eigenvalue $x^\top Ax$. Moreover, since $x \in \mathbb{S}^{d-1}$, we have $x^\top x = 1$, which implies x is a unit eigenvector of A.
Let $\lambda_1, \cdots, \lambda_d$ be the eigenvalues of A, and $x_1, \cdots, x_d$ be the corresponding orthonormal eigenvectors. Then, x is a critical point of f implies $x = x_i$ for some i.
Furthermore, $\operatorname{Hess} f(x)[u] \succeq 0$ implies for any $u \in T_x\mathbb{S}^{d-1}$, we have

$$
\begin{aligned}
0 &\le \langle u, \operatorname{Hess} f(x)[u]\rangle \\
&= \langle u, (I - xx^\top)Au - (x^\top Ax)u\rangle \\
&= u^\top (I - xx^\top)Au - (x^\top Ax)u^\top u \\
&= u^\top Au - \lambda_i \|u\|^2 .
\end{aligned}
$$

In particular, since $\{x_1, \ldots, x_d\}$ is an orthonormal basis of $\mathbb{R}^d$, then $x_j \in T_{x_i}\mathbb{S}^{d-1}$ for $j \neq i$, thus we can choose $u = x_j$ to get

$$
\begin{aligned}
0 &\le x_j^\top Ax_j - \lambda_i \|x_j\|^2 \\
&= \lambda_j - \lambda_i,
\end{aligned}
$$

which implies λ_i is the smallest eigenvalue of A, i.e., $x = x_i$ is globally optimal of f.

□

Problem 3 Score: ________. Geodesics on the sphere

Let $\mathbb{S}^{d-1}$ be the $(d-1)$-dimensional sphere embedded in $\mathbb{R}^d$ with the usual inner product. Let $(x, v) \in T\mathcal{M}$, consider

$$
c(t) = \cos(t)x + \sin(t)\frac{v}{\|v\|}.
$$

(1) Show that the curve $c(t)$ is a geodesic on $\mathbb{S}^{d-1}$.

Solution: (1) Since $\|c(t)\| = \cos^2(t) + \sin^2(t) = 1$, we have $c(t) \in \mathbb{S}^{d-1}$.
The velocity of $c(t)$ is given by

$$
\dot{c}(t) = -\sin(t)x + \cos(t)\frac{v}{\|v\|}.
$$

Thus at $t = 0$, we have

$$
\begin{aligned}
c(0) &= x \\
\dot{c}(0) &= -\sin(0)x + \cos(0)\frac{v}{\|v\|} \\
&= \frac{v}{\|v\|}.
\end{aligned}
$$

Then, the acceleration of $c(t)$ is given by

$$
\begin{aligned}
\ddot{c}(t) &= \frac{D}{dt}\dot{c}(t) \\
&= \operatorname{Proj}_{c(t)}\left(\frac{d}{dt}\dot{c}(t)\right)
\end{aligned}
$$

$$= \mathrm{Proj}_{c(t)}\left(-\cos(t)x - \sin(t)\frac{v}{\|v\|}\right).$$

At $t = 0$, we have

$$\begin{aligned}
\ddot{c}(0) &= \mathrm{Proj}_x\left(-x\right) \\
&= (I - xx^\top)(-x) \\
&= -x + x \\
&= 0.
\end{aligned}$$

□