Exercise 5

Due date:

Name : Vivi
Student ID : 24S153073
Grade : _________

Problem 1 Score: _________. RGD on product of spheres
Let $\mathcal{M} = \mathbb{S}^{m-1} \times \mathbb{S}^{n-1}$, which is an embedded submanifold of $\mathcal{E} = \mathbb{R}^m \times \mathbb{R}^n$ with its usual Euclidean structure. We turn $\mathcal{M}$ into a Riemannian manifold by using the Euclidean structure of the ambient space $\mathcal{E} = \mathbb{R}^m \times \mathbb{R}^n$.
Let $M \in \mathbb{R}^{m\times n}$ and

$$f : \mathcal{M} \to \mathbb{R}, \quad f(x, y) = x^\top My.$$

(1) Show that $\max\limits_{(x,y)\in\mathcal{M}} f(x, y) = \sigma_1(M)$, where $\sigma_1(M)$ is the largest singular value of M.

(2) Characterize the critical points of f on $\mathcal{M}$, and relate them to the eigenvectors of $M^\top M$ and $MM^\top$.

(3) Propose a retraction for $\mathcal{M}$.

(4) Write down the RGD algorithm for $-f$ on $\mathcal{M}$.

(5) Write code of the algorithm.

Solution: (1) For $(x, y) \in \mathcal{M}$, we have

$$\begin{aligned} x^\top My &\le \left\|x^\top My\right\|_2 \\ &\le \|x\|_2 \cdot \|M\|_2 \cdot \|y\|_2 \\ &= \sigma_1(M) \end{aligned}$$

When x, y are the singular vectors corresponding to $\sigma_1(M), x^\top My = \sigma_1(M)$.

(2) For $(x, y) \in \mathcal{M}$, we have

$$gradf(x, y) = ((I - xx^\top)My, (I - yy^\top)M^\top x).$$

If $gradf(x, y) = 0$, then

$$\begin{aligned} My &= (x^\top My)x \\ M^\top x &= (x^\top My)y \end{aligned}$$

Hence,

$$\begin{aligned} M^\top My &= (x^\top My)M^\top x = (x^\top My)^2 y \\ MM^\top x &= (x^\top My)My = (x^\top My)^2 x \end{aligned}$$

which means that x is an eigenvector of $MM^\top$ and y is an eigenvector of $M^\top M$, corresponding to the eigenvalue $(x^\top My)^2$.

(3) The retraction for $\mathcal{M}$ can be

$$R : T\mathcal{M} \to \mathcal{M}, \quad R_{(x,y)}(v, w) = \left(\frac{x + v}{\|x + v\|_2}, \frac{y + w}{\|y + w\|_2}\right).$$

□

Problem 2 Score: _________. RGD on Stiefel For $p \le n$, consider the Stiefel manifold

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

We endow $\mathcal{M}$ with the inner product $\langle X, Y\rangle = \mathrm{Tr}(X^\top Y)$. Let

$$f : \mathcal{M} \to \mathbb{R}, \quad f(X) = \mathrm{Tr}(X^\top AX),$$

where A is a real symmetric $n \times n$ matrix.

(1) Compute the orthogonal projector $Proj_X : \mathcal{E} \to T_X\mathcal{M}$.

(2) Given $X \in \mathcal{M}$ and $U \in \mathcal{E}$, give its time complexity to compute $Proj_X(U)$.

(3) Give an expression of the Riemannian gradient $gradf(X)$.

We want to solve

$$\min_{X \in \mathcal{M}} f(X),$$

which amounts to identifying a left invariant subspace of A.

Solution: (1) For $X \in \mathcal{M}$, we have

$$T_X\mathcal{M} = \{V \in \mathbb{R}^{n \times p} : X^\top V + V^\top X = 0\}, \quad N_X\mathcal{M} = \{XA : A \in Sym(p)\}.$$

Hence, for $U \in \mathbb{R}^{n \times p}$, we have $Proj_X(U) \in T_X\mathcal{M}$ and $U - Proj_X(U) \in N_X\mathcal{M}$, we have

$$\begin{aligned}
Proj_X(U)^\top X + X^\top Proj_X(U) &= (U - XA)^\top X + X^\top(U - XA) \\
&= U^\top X - A^\top X^\top X + X^\top U - X^\top XA \\
&= U^\top X + X^\top U - 2A \\
&= 0
\end{aligned}$$

that is, $A = \frac{U^\top X + X^\top U}{2}$, therefore $Proj_X(U) = U - XA = U - X\frac{U^\top X + X^\top U}{2}$.

(2) The time complexity to compute $Proj_X(U)$ is $O(np^2)$.

(3) f can be smoothly extended by

$$\bar{f} : \mathbb{R}^{n \times p} \to \mathbb{R}, \quad \bar{f}(X) = \mathrm{Tr}(X^\top AX).$$

The gradient of $\bar{f}$, $grad\bar{f}(X) = 2AX$. Hence, the Riemannian gradient of f is

$$\begin{aligned}
gradf(X) &= Proj_X(grad\bar{f}(X)) \\
&= Proj_X(2AX) \\
&= 2Proj_X(AX) \\
&= 2(AX - X\frac{X^\top AX + X^\top AX}{2}) \\
&= 2(AX - XX^\top AX)
\end{aligned}$$

□

Problem 3 Score: _________. PL-condition, sufficient decrease and linear convergence

Let $\mathcal{M}$ be a Riemannian manifold and $f : \mathcal{M} \to \mathbb{R}$ be a smooth function sutisfying the PL-condition:

$$\exists \mu > 0 s.t. \left\| gradf(x) \right\|_x^2 \geq 2\mu(f(x) - f^*), \quad \forall x \in \mathcal{M},$$

where $f^* = \min_{x \in \mathcal{M}} f(x)$.

Consider an iterative algorithm $\mathcal{A}$ with iterates $\{x_k\}$ satisfying the sufficient decrease condition:

$$\exists c > 0 s.t. f(x_{k+1}) - f(x_k) \leq -c \left\| gradf(x_k) \right\|_{x_k}^2, \quad \forall k = 0, 1, 2, \cdots$$

Backtracking line-search satisfies sufficient decrease(assuming Lipschitz conditions on f)

(1) Show that algorithm $\mathcal{A}$ converges at a linear rate:

$$f(x_{k+1}) - f^* \leq (1 - 2\mu c)(f(x_k) - f^*), \quad \forall k = 0, 1, 2, \cdots$$

(2) Show that if $f : \mathcal{M} \to \mathbb{R}$ satisfies the PL-condition, then all critical points of f are global minimizers.

(3) If $\mathcal{M}$ is a sphere, can a non constant function $f : \mathcal{M} \to \mathbb{R}$ satisfy the PL-condition? What about if $\mathcal{M}$ is a compact Riemannian manifold ?

(4) Let $\mathcal{M} = \mathbb{R}^d$ endowed with the standard inner product ($\mathcal{M}$ is a Euclidean space). Show that if f is differentiable and μ-strongly convex, i.e.,

$$f(y) \geq f(x) + \langle gradf(x), y - x \rangle + \frac{\mu}{2} \left\| y - x \right\|_2^2, \quad \forall x, y \in \mathcal{M},$$

then f satisfies the PL-condition with μ.

Solution: (1) We have

$$\begin{aligned} 2\mu c(f(x_k) - f^*) &\leq c \left\| gradf(x_k) \right\|_{x_k}^2 \\ &\leq f(x_k) - f(x_{k+1}) \\ &= (f(x_k) - f^*) - (f(x_{k+1}) - f^*) \end{aligned}$$

which means that

$$f(x_{k+1}) - f^* \leq (1 - 2\mu c)(f(x_k) - f^*).$$

(2) If $x \in \mathcal{M}$ is a critical point of f, then $gradf(x) = 0$. By the PL-condition, we have

$$0 \geq 2\mu(f(x) - f^*),$$

which means that $f(x) = f^*$.

(3) Suppose that $f : \mathcal{M} \to \mathbb{R}$ satisfies the PL-condition. Then, for $x_{max} \in \mathrm{argmax}_{x \in \mathcal{M}} f(x)$, we have

$$0 \geq 2\mu(f(x_{max}) - f^*),$$

which means that $f(x_{max}) = f^*$. Hence, f is constant.

(4) Fix $x \in \mathcal{M}$, we define

$$g(y) = f(x) + \langle gradf(x), y - x \rangle + \frac{\mu}{2} \left\| y - x \right\|_2^2.$$

Then, we have

$$Dg(y)[v] = \langle gradf(x) + \mu(y - x), v \rangle, \quad v \in \mathbb{R}^d.$$

The only critical point of g is $y^* = x - \frac{1}{\mu} gradf(x)$. Hence

$$f(y) \geq f(x) + \langle gradf(x), y - x \rangle + \frac{\mu}{2} \left\| y - x \right\|_2^2 \geq f(y^*) = f(x) - \frac{1}{2\mu} \left\| gradf(x) \right\|_2^2.$$

That means

$$\left\| gradf(x) \right\|_2^2 \geq 2\mu(f(x) - f(y)), \quad \forall x, y \in \mathcal{M}.$$

In particular, for $y = \mathrm{argmin}_{x \in \mathcal{M}} f(x)$, we have

$$\left\| gradf(x) \right\|_2^2 \geq 2\mu(f(x) - f^*), \quad \forall x \in \mathcal{M}.$$

□