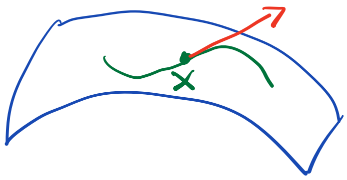
Setup

- $\mathcal{E}$ is a linear space with inner product $\langle , \rangle$
- $M \subseteq \mathcal{E}$ is a submanifold in $\mathcal{E}$
- The general definition of tangent space of a manifold $M \subseteq \mathcal{E}$

$$T_x M = \{ c'(0) \mid c : \underbrace{(-\varepsilon, \varepsilon)}_{\text{some interval in } \mathbb{R}} \to M \text{ is smooth } \& \ c(0) = x \}$$



$c(t)$ is a curve on M

$c'(t)$ is tangent to the curve

Theorem If $M \subseteq \mathcal{E}$ is describle by $h(x) = 0$, then

$$\mathrm{Ker}[Dh(x)] = T_x M$$

Any element in $T_x M$ is a tangent vector at x

- Tangent Bundle $TM = \{ V_x \in T_x M : \forall x \in M \} = \bigcup_{x \in M} T_x M$

subscript x deotes it's a tangent vector at x

For convenience, we sometimes ignore the subscript.

- A vector field on M is a map

$$V : M \longrightarrow TM$$

s.t. $V(x) \in T_x M$ for any $x \in M$

- A metric is a choice of inner product $\langle \cdot , \cdot \rangle_x$

for $T_x M$ at each $x \in M$.

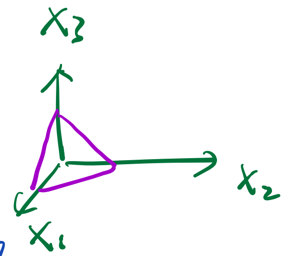
- A metric is Riemannian if $\langle \cdot , \cdot \rangle_x$ is smooth w.r.t. x

  i.e. $\langle V(x), W(x) \rangle_x$ is smooth for smooth vector fields V, W.

- Let $g_x(\cdot , \cdot) = \langle \cdot , \cdot \rangle_x$ be a Riemannian metric

  (M, g) is a Riemannian manifold.

Δ^2_+



Example:

$$M = \Delta^{n-1}_+ = \{ x \in \mathbb{R}^n : x_i > 0 , \sum_{i=1}^n x_i = 1 \}$$

relative interior of the simplex or probability manifold

M is an embedded manifold of $\mathbb{R}^n$

$$T_x M = \{ v \in \mathbb{R}^n : \sum_{i=1}^n v_i = 0 \}$$

The Fisher-Rao metric (M, g)

$$g_x(u, v) = \sum_{i=1}^n \frac{u_i v_i}{x_i}$$

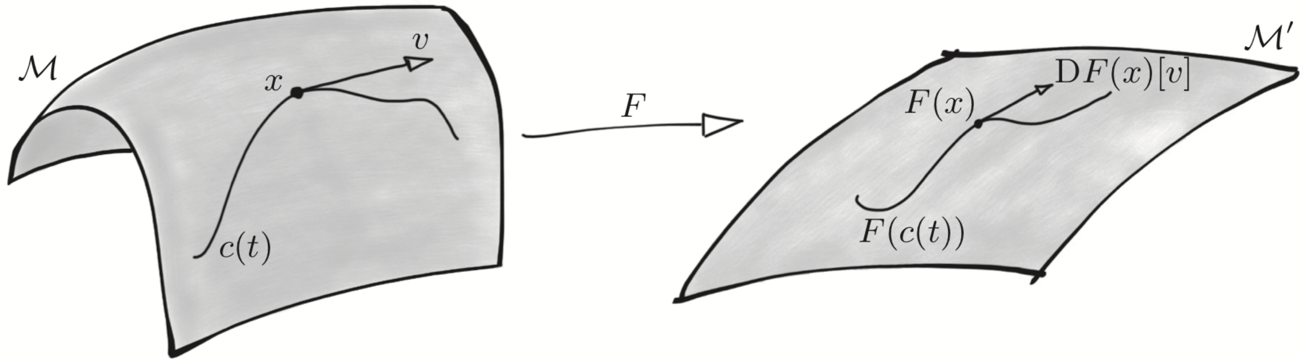
- $F : M \to M'$ is smooth map where M, M' are manifolds

The differential is the linear map

$DF(x) : T_x M \to T_{F(x)} M'$ defined by

$$DF(x)[v] = \frac{d}{dt} F(c(t)) \Big|_{t=0} = (F \circ c)'(0)$$

where $c(t)$ is a smooth curve on M $\begin{cases} c(0) = x \\ c'(0) = v \end{cases}$



Example $\quad f : \mathcal{M} \to \mathbb{R}$

$$Df(x) : T_x\mathcal{M} \to \mathbb{R}$$

$$v \longmapsto Df(x)[v]$$

$$= \frac{d}{dt} f(c(t)) \Big|_{t=0}$$

not easy to compute

- Let $\mathcal{M} \subset \mathcal{E}$ and $\mathcal{M}' \subset \mathcal{E}'$ be embedded submanifolds in $\mathcal{E}$ and $\mathcal{E}'$

Then a smooth map $F : \mathcal{M} \to \mathcal{M}'$ has a smooth extension $\bar{F} : \mathcal{E} \to \mathcal{E}'$

Notice that $F(c(t)) = \bar{F}(c(t))$ because $c(t) \in \mathcal{M}$.

$$\frac{d}{dt} \bar{F}(c(t)) \Big|_{t=0} = D\bar{F}(c(0))[c'(0)] = D\bar{F}(x)[v]$$

$$\|$$

$$\frac{d}{dt} F(c(t)) \Big|_{t=0} = DF(x)[v]$$

$$\Rightarrow \quad DF(x)[v] = D\bar{F}(x)[v]$$

Example $\quad St(n,p) = \{ X \in \mathbb{R}^{n \times p},\ X^T X = I_p \} \quad p \leq n$

$$f : St(n,p) \longrightarrow \mathbb{R} \qquad \langle U, V\rangle = tr(U^T V)$$

$$\boxed{A^T = A} \qquad X \longmapsto tr(X^T A X) = \langle X, AX\rangle$$

$$\langle X, Y\rangle = \sum_{i=1}^{n}\sum_{j=1}^{p} X_{ij} Y_{ij}$$

can be extended to

$$\bar{f} : \mathbb{R}^{n\times p} \longrightarrow \mathbb{R}$$

$$X \longmapsto tr(X^T A X) = \langle X, AX\rangle$$

$$\forall V \in \mathbb{R}^{n\times p}, \quad D\bar{f}(X)[V] = \lim_{t\to 0} \frac{\bar{f}(X+tV) - \bar{f}(X)}{t}$$

$$= \lim_{t\to 0} \frac{\langle X+tV, A(X+tV)\rangle - \langle X, AX\rangle}{t}$$

$$= \lim_{t\to 0} \frac{t\langle V, AX\rangle + t\langle X, AV\rangle + t^2\langle V, AV\rangle}{t}$$

$$= \langle 2AX, V\rangle$$

$\forall X \in St(n,p)$

$$\forall V \in T_X St(n,p) = \{ V \in \mathbb{R}^{n\times p} : X^T V + V^T X = 0\} \subseteq \mathbb{R}^{n\times p}$$

$$Df(X)[V] = D\bar{f}(X)[V] = \langle 2AX, V\rangle$$

- Riemannian Gradient

For a smooth $f : \mathcal{M} \to \mathbb{R}$ on a Riemannian manifold $(\mathcal{M}, g)$ the Riemannian gradient of f is the vector field grad f defined by $g_x(\text{grad} f(x), V_x) = Df(x)[V_x]$, $\forall x \in \mathcal{M}$ $\forall V_x \in T_x\mathcal{M}$

Remark: $\operatorname{grad} f(x)$ is a tangent vector in $T_x M$

Example:
$$f: St(n,p) \longrightarrow \mathbb{R}$$
$$X \longmapsto tr(X^T A X) = \langle X, AX \rangle$$

$$\bar{f}: \mathbb{R}^{n\times p} \longrightarrow \mathbb{R}$$
$$X \longmapsto tr(X^T A X) = \langle X, AX \rangle$$

$$Df(X)[V] = D\bar{f}(x)[V] = \langle 2AX, V \rangle$$

① Euclidean Gradient of $\bar{f}$ in $\mathbb{R}^{n\times p}$ is defined by

$$\langle \nabla \bar{f}(X), V \rangle = D\bar{f}(x)[V], \quad \forall V \in \mathbb{R}^{n\times p}$$

$$\Rightarrow \nabla \bar{f}(x) = 2AX$$

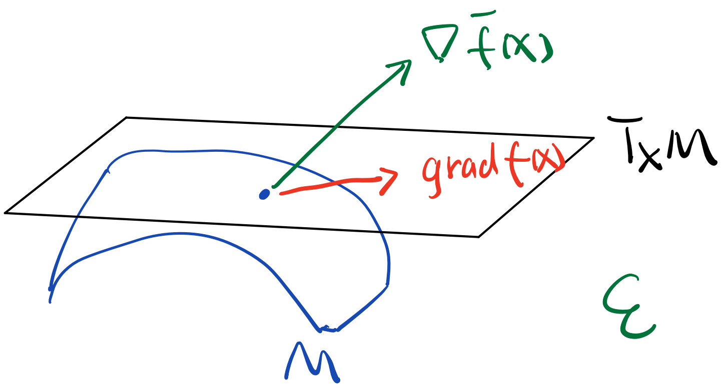
② Riemannian Gradient of f on (M, g) is defined by

$$g_x(\operatorname{grad} f(x), V) = Df(X)[V], \quad \forall V \in T_x M \subseteq \mathbb{R}^{n\times p}$$

Suppose we choose $g_x(U,V) = \langle U, V \rangle$, then

$$\langle \operatorname{grad} f(x), V \rangle = \langle \nabla \bar{f}(X), V \rangle, \quad \forall V \in T_x M$$

$$\Rightarrow \langle \operatorname{grad} f(x) - \nabla \bar{f}(x), V \rangle = 0, \quad \forall V \in T_x M$$



$\Rightarrow$ $\operatorname{grad} f(x)$ is Euclidean Projection of $\nabla \bar{f}(X)$ onto $T_x M$

③ $T_X St(n,p) = \{ V \in \mathbb{R}^{n\times p} : X^T V + V^T X = 0 \} = S_1$

$X_\perp \in \mathbb{R}^{n\times(n-p)}$ is a matrix s.t.

Col space of $X_\perp$ is orthogonal complement to Col space of $X \in \mathbb{R}^{n\times p}$

$T_X St(n,p) = \{ X\Omega + X_\perp K : \Omega^T = -\Omega \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p} \} = S_2$

Pick $X_\perp$ s.t. $X_\perp$ has orthonormal columns

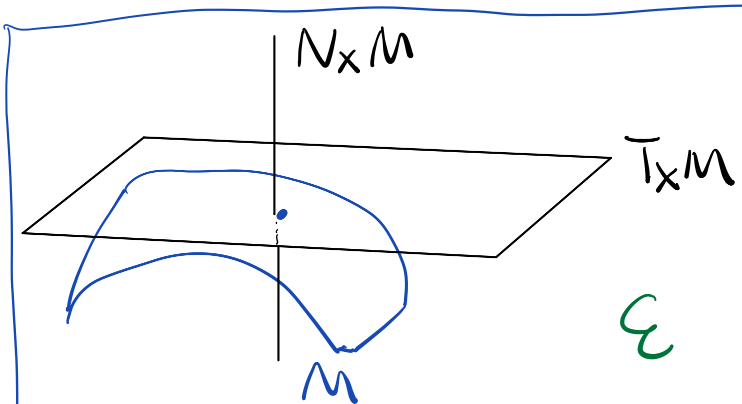
The matrix $[X \ X_\perp] \in \mathbb{R}^{n\times n}$ has orthonormal columns

This can be proven by counting dimensions.

1) $S_2 \subseteq S_1$ because $(X\Omega + X_\perp K)^T X + X^T (X\Omega + X_\perp K)$

$$= \Omega^T X^T X + K^T X_\perp^T X + X^T X \Omega + X^T X_\perp K = 0$$

2) S_2 has the same dim as $S_1 \Rightarrow S_2 = S_1$



Normal Space $N_X M$ is the orthogonal complement of $T_X M$ in $\mathcal{E}$

③ $N_X St(n,p) = \{ XS : S^T = S, S \in \mathbb{R}^{p\times p} \}$

$$\langle XS, X\Omega + X_\perp K \rangle = tr(S X^T (X\Omega + X_\perp K))$$

$$= tr(S\Omega)$$

$$= \langle S, \Omega \rangle = 0$$

because $S^T = S$, $\Omega^T = -\Omega$

④ Projection of onto $N_X St(n,p)$ is $P_X^{\perp}$

Projection of onto $T_X St(n,p)$ is P_X

Then $P_X(Y) = Y - P_X^{\perp}(Y) = (I - P_X^{\perp})Y$

$$P_X^{\perp}(Y) = X\,\mathrm{sym}(X^T Y) = X\,\frac{X^T Y + Y^T X}{2}$$

$$\mathrm{sym}(B) = \frac{B^T + B}{2}$$

$$P_X(Y) = (I - XX^T)Y + \mathrm{skew}(X^T Y)$$

$$\mathrm{skew}(B) = \frac{1}{2}(B - B^T)$$

$$\Rightarrow \mathrm{grad} f(x) = P_X[\nabla \bar{f}(x)]$$

$$= (I - XX^T)\nabla \bar{f}(x) + \mathrm{skew}(X^T \nabla \bar{f}(x))$$

Special Example :

① $p=1$, $St(n,p) = S^{n-1}$

$$P_X(Y) = (I - XX^T)Y + \mathrm{skew}(X^T Y) = (I - XX^T)Y$$

$$\mathrm{grad} f(x) = (I - XX^T)\nabla \bar{f}(x)$$

② $p=n$, $St(n,p) = O(n) = \{X \in \mathbb{R}^{n\times n} : X^T X = I\}$

$$f(x) = \frac{1}{2}\mathrm{tr}(X^T A X), \quad A^T = A$$

$$\nabla \bar{f}(x) = AX \Rightarrow \mathrm{skew}(X^T \nabla \bar{f}(x)) = \mathrm{skew}(X^T A X) = 0$$

$$\Rightarrow \mathrm{grad} f(x) = (I - XX^T)AX$$

Example ① Regard S^{n-1} as a Riemannian submanifold of $\mathbb{R}^n$

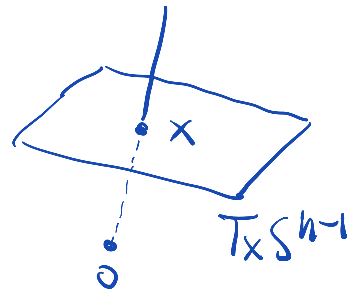
$f(x) = \frac{1}{2}x^TAx$ defined on S^{n-1}

$\bar{f}(x) = \frac{1}{2}x^TAx$ defined on $\mathbb{R}^n$

$\Big\}\ A^T = A$

$\nabla \bar{f}(x) = Ax$

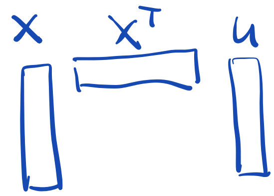
Projection of u onto T_xS^{n-1}



$$P_{T_xS^{n-1}}(u) = u - \frac{\langle x, u\rangle}{\langle x, x\rangle}x \quad , x \in S^{n-1}$$

$$= u - \langle x, u\rangle x$$

$$= (I - xx^T)u$$



$$\operatorname{grad} f(x) = P_{T_xS^{n-1}}(\nabla \bar{f}(x))$$

$$= (I - xx^T)Ax = Ax - (x^TAx)x$$

Remark: $\operatorname{grad} f(x) = 0 \iff Ax = \underbrace{(x^TAx)}_{\text{scalar}}\, x$

$\nabla \bar{f}(x) = Ax$

So critical point on $S^{n-1} \iff$ eigenvectors

② $M = \Delta_+^{n-1} = \{x \in \mathbb{R}^n : x_i > 0,\ \sum_{i=1}^n x_i = 1\}$

relative interior of the simplex or probability manifold

M is an embedded manifold of $\mathbb{R}^n$

$T_xM = \{v \in \mathbb{R}^n : \sum_{i=1}^n v_i = 0\}$

The Fisher-Rao metric

$$g_x(u,v) = \sum_{i=1}^{n} \frac{u_i v_i}{x_i}$$

$$f : M \to \mathbb{R}$$

$$\bar{f} : \mathbb{R}^n \to \mathbb{R} \qquad \nabla \bar{f}(x)$$

$$g_x(\operatorname{grad} f(x), v) = Df(x)[v] = D\bar{f}(x)[v]$$
$$= \langle \nabla \bar{f}(x), v \rangle$$
$$= v^T \cdot \nabla \bar{f}(x)$$

To find an expression

Let $\operatorname{grad} f(x) = u \in T_x M$

$\frac{u}{x}$ denotes the vector $\left[\frac{u_1}{x_1} \cdots \frac{u_n}{x_n}\right]^T$

$$v^T \cdot \frac{\operatorname{grad} f(x)}{x} = v^T \cdot \nabla \bar{f}(x) \qquad \forall v \in T_x M$$

Given $w \in \mathbb{R}^n$, solve $u \in T_x M$ satisfying

$$v^T u = v^T w, \quad \forall v \in T_x M$$

$$\Rightarrow v^T (u - w) = 0$$

$$\Rightarrow P_{T_x M}(u - w) = 0$$

$$\Rightarrow u - w \ // \begin{bmatrix} | \\ \vdots \\ | \end{bmatrix}$$

$$\Rightarrow u = w + \lambda \mathbb{1}, \quad \lambda \in \mathbb{R}$$

$$\Rightarrow \frac{\operatorname{grad} f(x)}{x} = \nabla \bar{f}(x) + \lambda \mathbb{1}$$

$$\Rightarrow \quad \operatorname{grad} f(x) = \underline{x \circ} \nabla \bar{f}(x) + \lambda x$$

entry wise product

$$\operatorname{grad} f(x) \in T_x M \Rightarrow \sum_{i=1}^{n} \left(x_i [\nabla \bar{f}(x)]_i + \lambda x_i \right) = 0$$

$$\Rightarrow \lambda = -\sum_{i=1}^{n} x_i [\nabla \bar{f}(x)]_i$$

$$\Rightarrow \quad \operatorname{grad} f(x) = x \circ \nabla \bar{f}(x) - \langle x, \nabla \bar{f}(x) \rangle x$$

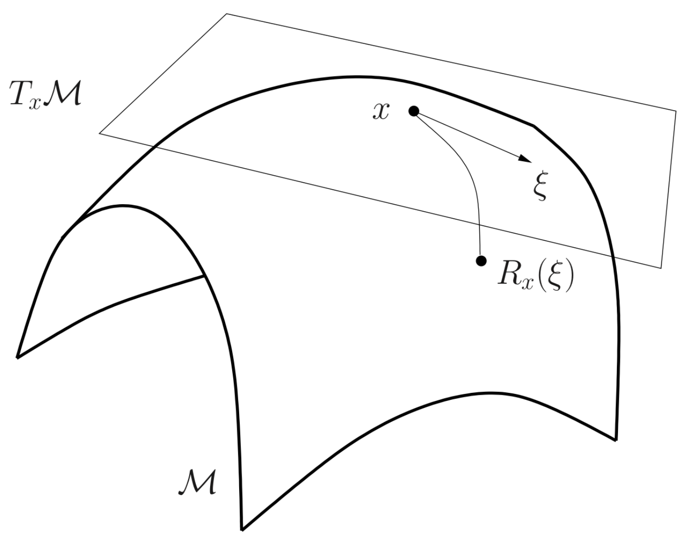
A **Retraction** on M is a smooth map

$$R: TM \longrightarrow M$$
$$v_x \longmapsto R_x(v)$$

such that each curve $c(t) = R_x(tv)$ satisfies

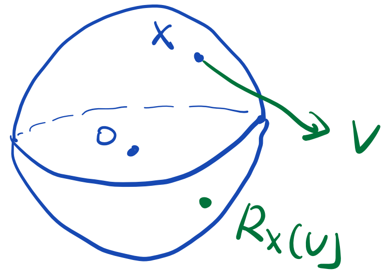
$$\begin{cases} c(0) = x \\ c'(0) = v \end{cases}$$

$$R: TM \to M$$
$$v_x \longmapsto R_x(v_x)$$



$$c(t) = R_x(t\xi)$$
$$c(0) = x$$
$$c'(0) = \xi$$

Examples of Retraction on S^{n-1}



① $R_x(v) = \dfrac{x+v}{\|x+v\|} = \dfrac{x+v}{\sqrt{\|x+v\|^2}} = \dfrac{x+v}{\sqrt{1+\|v\|^2}}$

② $R_x(v) = \cos(\|v\|)\,x + \dfrac{\sin(\|v\|)}{\|v\|}\,v$

$c(t) = R_x(tv)$ is the great circle passing x with $c'(0)=v$.

Riemannian Gradient descent on (M, g)

$$\min_{x\in M} f(x)$$

$$x_{k+1} = R_{x_k}[-\eta \operatorname{grad} f(x)]$$

Example:

M = sphere S^{n-1}, g is $\langle\, , \rangle$, $f(x) = \frac{1}{2}x^T A x$

$$R_x[v] = \frac{x+v}{\|x+v\|}$$

$$\begin{cases} \tilde{x}_{k+1} = x_k - \eta\,(I - x_k^T x_k)\,A x_k \\ x_{k+1} = \dfrac{\tilde{x}_{k+1}}{\|\tilde{x}_{k+1}\|} \end{cases}$$