Given a Riemannian manifold (M, g) with $M \subseteq \mathcal{E}$,

and $f: M \to \mathbb{R}$ can be extended as

$\bar{f}: \mathcal{E} \to \mathbb{R}$ s.t. $\bar{f}|_M = f$

Riemannian Gradient Descent of minimizing $f(x)$ over M

① Riemannian Gradient of $f(x)$ on (M, g) $\operatorname{grad} f(x)$

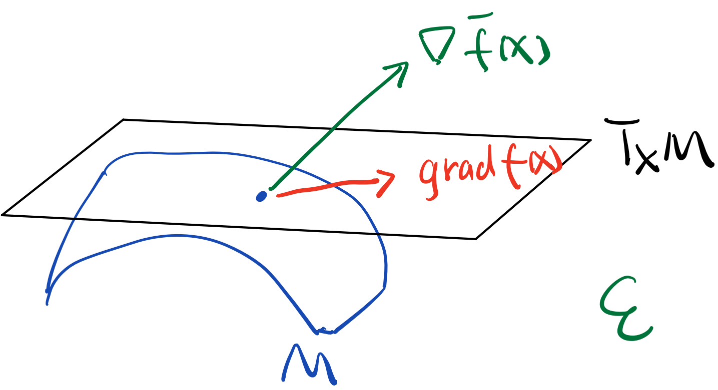
$$\forall V \in T_x M, \; g_x(\operatorname{grad} f(x), V) = Df(x)[V]$$
$$= D\bar{f}(x)[V]$$
$$= \langle \nabla \bar{f}(x), V \rangle$$

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

Example: If g_x is $\langle , \rangle$, then

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

$$\Rightarrow \operatorname{grad} f(x) = P_{T_x M}(\nabla \bar{f}(x))$$



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

$\mathcal{E} = \mathbb{R}^{n \times p}$, $g(U, V) = \langle U, V \rangle$

$$P_{T_x M}(Y) = (I - XX^T)Y + \operatorname{skew}(X^T Y)$$

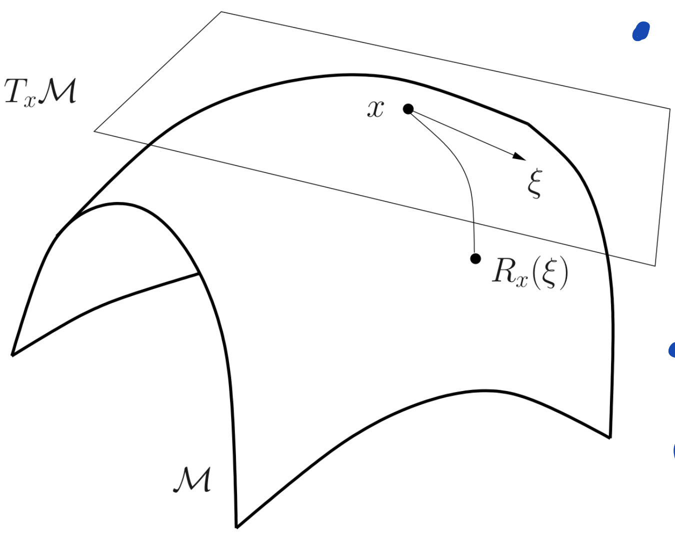
$$\operatorname{skew}(B) = \frac{1}{2}(B - B^T)$$

② Retraction at $x \in \mathcal{M}$ is a map

$$R_x : T_x\mathcal{M} \to \mathcal{M}$$
$$\xi \mapsto R_x(\xi)$$

we perceive ξ as $x+\xi \in \mathcal{E}$

since we identify $0 \in T_x\mathcal{M}$ with $x \in \mathcal{M} \subseteq \mathcal{E}$



- $\forall t \in (-\varepsilon, \varepsilon) \subset \mathbb{R}$, $t\xi$ (perceived as $x+t\xi$) is a straight line segment on $T_x\mathcal{M}$
- R_x maps the line to a curve $c(t) = R_x(t\xi) \in \mathcal{M}$, $\forall t$

Definition/Requirement of a Retraction at x:

1) $c(0) = x$ (curve passes x)
2) $c'(0) = \xi$ (velocity is the input)

which can be equivalently written as

1) $R_x(0) = x$
   zero vector in $T_x\mathcal{M}$, perceived as $x \in \mathcal{E}$
2) $DR_x(0)$ is identity map from $T_x\mathcal{M}$ to $T_x\mathcal{M}$

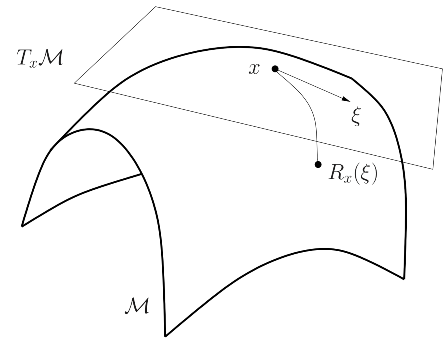
$R_x : T_xM \longrightarrow M$ is a map

$\xi \longmapsto R_x(\xi)$

$\mathcal{T} = T_xM$ is also an embedded manifold in $\mathcal{E}$

$DR_x(\xi) : T_\xi \mathcal{T} \longrightarrow T_{R_x(\xi)} M$ is the differential map at ξ

$V \longmapsto DR_x(\xi)[V]$



Since $\mathcal{T} = T_xM \subseteq \mathcal{E}$, $M \subseteq \mathcal{E}$,

R_x can be extended to

$\bar{R}_x : \mathcal{E} \longrightarrow \mathcal{E}$ s.t. $\bar{R}_x |_{\mathcal{T}} = R_x$

and $DR_x(\xi)[V] = D\bar{R}_x(\xi)[V]$, $\forall V \in T_\xi \mathcal{T}$

Now set $\xi = 0$, we consider $DR_x(0)$

$DR_x(0) : T_0\mathcal{T} \longrightarrow T_{R_x(0)} M$ is the map

$DR_x(0) : T_xM \longrightarrow T_xM$ if we identify $T_0\mathcal{T}$ with $\mathcal{T}$

tangent plane of a plane is the same plane.

$V \longmapsto DR_x(0)[V]$

$\qquad\qquad\qquad\quad \| $

$\qquad\qquad\qquad\quad V$

requirement of a retraction

Let's compute $DR_x(0)[V]$ by definition, using the curve

$c(t) = R_x(tV)$

$\forall v \in T_x M \simeq T_\xi \Upsilon$, $\quad DR_x(\xi)[v] = D\bar{R}_x(\xi)[v]$

$$= \lim_{t \to 0} \frac{\bar{R}_x(\xi + tv) - \bar{R}_x(\xi)}{t}$$

At $\xi = 0$, $\forall v \in T_x M \simeq T_0 \Upsilon$,

$$DR_x(0)[v] = D\bar{R}_x(0)[v]$$

$$= \lim_{t \to 0} \frac{\bar{R}_x(0 + tv) - \bar{R}_x(0)}{t}$$

$$= \lim_{t \to 0} \frac{R_x(tv) - R_x(0)}{t}$$

$$= \lim_{t \to 0} \frac{c(t) - c(0)}{t} = c'(0)$$

So

$$\begin{cases} c(t) \text{ satisfying} \\ c(0) = x \\ c'(0) = v \end{cases} \iff \begin{cases} R_x \text{ satisfying} \\ R_x(0) = x \\ DR_x(0)[v] = v \end{cases}$$

③ <u>Def</u> (Projection in $\mathcal{E}$)

For fixed $y \in \mathcal{E}$, Projection onto M is defined as

$$P_M(y) = \underset{x \in M}{\operatorname{argmin}} \|x - y\|^2 = \langle x-y, x-y \rangle$$

Remark: $P_M(y)$ can be single point, multiple points or $P_M(y)$ may not exist

<u>Theorem</u> For an embedded manifold $M \subseteq \mathcal{E}$,

$R_x : T_xM \to M$ defined by

$\xi \mapsto P_M(x+\xi)$ is a retraction

Remark:

Ⅰ This projection-defined retraction is the most intuitive choice

Ⅱ A retraction has no dependence on the metric g of a Riemannian manifold

Ⅲ When g is not $\langle , \rangle$ in $\mathcal{E}$, grad $f(x)$ is NOT projection of $\nabla f(x)$ onto T_xM

But $P_M(x+\xi)$ is still a retraction

Ⅳ For a manifold M, a retraction is NOT unique.

④ <u>Second Order Retraction</u>

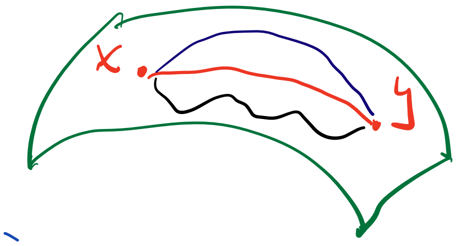
Retraction $\begin{cases} c(0) = x \\ c'(0) = \xi \end{cases}$

<u>Def</u> A Retraction R_x is second order if

$c(t) = R_x(t\xi)$ satisfies $c''(0) = 0$.

<u>Theorem</u> $P_M(x+\xi)$ is a second order retraction

⑤ Exponential Map



<u>Rough Definition of Geodesic</u>:

a geodesic connecting x & y on M is a curve on M connectiong x & y with shortest length

Example: Great Circle passing x, y on Sphere

<u>Def</u> Exponential Map $\mathrm{Exp}(x): T_xM \to M$

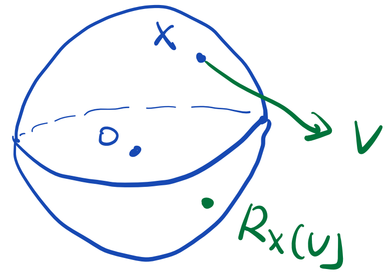
satisfies that $\begin{cases} c(t) = \mathrm{Exp}(x)[t\xi] \text{ is a geodesic} \\ c(0) = x, \quad c'(0) = \xi. \end{cases}$

<u>Fact</u>: 1) $\mathrm{Exp}(x)$ is a second-order retraction

2) A second order retraction is usually NOT $\mathrm{Exp}(x)$

3) $P_M(x+\xi)$ is often easier than $\mathrm{Exp}(x)[\xi]$

Examples of Retraction on S^{n-1}



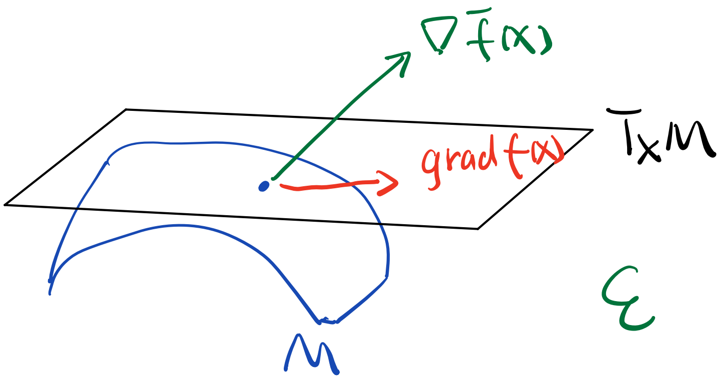
1) $R_x(v) = \dfrac{x+v}{\|x+v\|} = \dfrac{x+v}{\sqrt{\|x+v\|^2}}$

$$= \frac{x+v}{\sqrt{1+\|v\|^2}}$$ is the projection thus a second order retraction

2) $\mathrm{Exp}[x](v) = \cos(\|v\|)x + \dfrac{\sin(\|v\|)}{\|v\|} v$

$c(t) = \mathrm{Exp}[x](cv)$ is the great circle passing x with $c'(0) = v$.

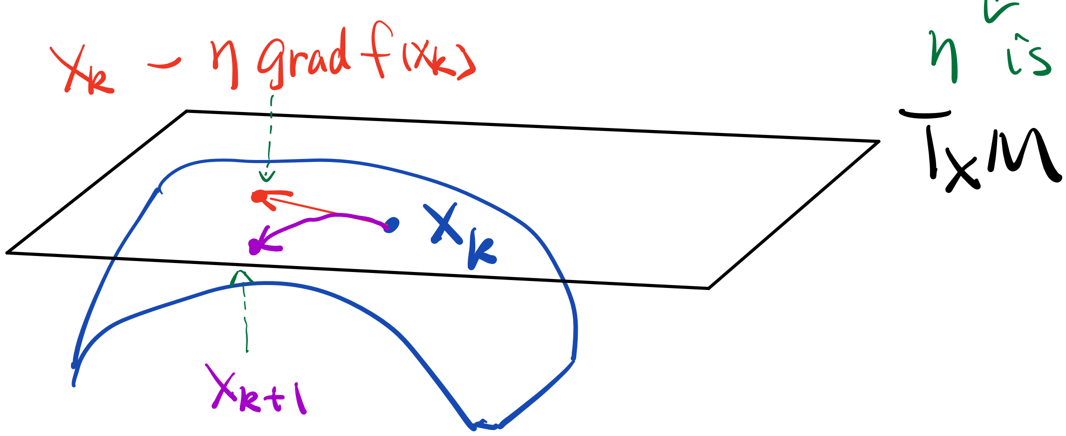
⑥ What is Riemannian Gradient Descent (RGD)



ε

Version I: $x_{k+1} = \mathrm{Exp}(x_k)[-\eta_k \operatorname{grad} f(x_k)]$

η is step size



Version II: $x_{k+1} = R_{x_k}[-\eta_k \operatorname{grad} f(x_k)]$

If R_x is projection-defined, then

$$x_{k+1} = P_M(x_k - \eta_k \operatorname{grad} f(x_k))$$