Examples and Summary of Riemannian Optimization

Example: $M = \mathbb{R}^n$ is a manifold

$\forall x \in M$, $T_xM = \mathbb{R}^n$

$M \subseteq \mathcal{E} = \mathbb{R}^n$

Given $f: M \to \mathbb{R}$, we still use $\bar{f}: \mathbb{R}^n \to \mathbb{R}$

Let g be a metric and $g_x(,)$ may not be $\langle , \rangle$

- Then the Riemannian Gradient of f on (M, g) is

$\forall v \in T_xM = \mathbb{R}^n$, $g_x(\text{grad} f(x), v) = Df(x)[v] = D\bar{f}(x)[v] = \langle \nabla \bar{f}(x), v \rangle$

- Retraction R_x can be taken as

$$R_x(v) = x + v$$

① If $g_x(,)$ is $\langle , \rangle$, then $\text{grad} f(x) \in T_xM = \mathbb{R}^n$

satisfies $\langle \text{grad} f(x), v \rangle = \langle \nabla \bar{f}(x), v \rangle$, $\forall v \in T_xM = \mathbb{R}^n$

$\Rightarrow \text{grad} f(x) = \nabla \bar{f}(x)$ which is $\nabla f(x)$

Riemannian Gradient Descent is

$$x_{k+1} = R_{x_k}(-\eta \, \text{grad} f(x_k))$$

$$= x_k - \eta \nabla f(x_k)$$

② If $g_x(u, v) = v^T G u = \langle Gu, v \rangle$

$$G \in \mathbb{R}^{n\times n}, \quad G^T = G, \quad G > 0$$

then $g_x(\text{grad} f(x), v) = \langle \nabla \bar{f}(x), v\rangle$

$$\Rightarrow \langle G \,\text{grad} f(x), v\rangle = \langle \nabla \bar{f}(x), v\rangle \quad \forall v \in \mathbb{R}^n$$

$$\Rightarrow \langle G \,\text{grad} f(x) - \nabla \bar{f}(x), v\rangle = 0$$

$$\Rightarrow \text{grad} f(x) = G^{-1} \nabla \bar{f}(x) = G^{-1} \nabla f(x)$$

Riemannian Gradient Descent is

$$x_{k+1} = R_{x_k}(-\eta \,\text{grad} f(x_k))$$

$$= x_k - \eta \underbrace{G^{-1} \nabla f(x_k)}_{\text{preconditioned gradient}}$$

$-G^{-1}\nabla f(x)$ is a descent direction because

$$G > 0 \Rightarrow G^{-1} > 0 \Rightarrow \langle \nabla f(x), G^{-1} \nabla f(x)\rangle = \nabla f(x)^T G^{-1} \nabla f(x) > 0$$

③ If $g_x(u, v) = v^T G(x) u = \langle G(x) u, v\rangle$

with $G(x) = \nabla^2 f(x)$, then

Riemannian Gradient Descent is Newton's Method

$$x_{k+1} = R_{x_k}(-\eta \,\text{grad} f(x_k))$$

$$= x_k - \eta [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Previously, we considered an embedded submanifold M in $\mathcal{E}$

- $M \subset \mathcal{E}$
- $f: M \to \mathbb{R}$ is extended to $\bar{f}: \mathcal{E} \to \mathbb{R}$
- T_xM is a subspace in $\mathcal{E}$
- $g_x(\operatorname{grad} f(x), v) = \langle \nabla \bar{f}(x), v \rangle$

We extend it to an embedded submanifold M in $(\bar{M}, \bar{g})$

where $\bar{M}$ is a manifold with a metric $\bar{g}$.

- $M \subset \bar{M}$ example: $\bar{M}$ is sphere, M is a circle on $\bar{M}$
- $f: M \to \mathbb{R}$ is extended to $\bar{f}: \bar{M} \to \mathbb{R}$
- T_xM is a subspace of $T_x\bar{M}$, $\forall x \in M \subset \bar{M}$
- We can define the metric of M as the one induced by $\bar{g}$: if $v \in T_xM$, then $v \in T_x\bar{M}$

  $$g_x(v, w) = \bar{g}_x(v, w), \ \forall v, w \in T_xM$$

(M, g) is called an embedded Riemannian submanifold of $(\bar{M}, \bar{g})$

- $\bar{f}: \bar{M} \to \mathbb{R}$ has a Riemannian Gradient $\operatorname{grad} \bar{f}$
- $f: M \to \mathbb{R}$, its Riemannian Gradient can be computed by

  $$g_x(\underbrace{\operatorname{grad} f(x)}_{\in T_xM}, v) = \bar{g}_x(\underbrace{\operatorname{grad} \bar{f}(x)}_{T_x\bar{M}}, v), \ \forall v \in T_xM$$

Example: $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $A^T = A$, $A > 0$

$$f(x) = \frac{1}{2} x^T A x + \frac{\beta}{4} \sum_{i=1}^{n} x_i^4 \quad , \beta \geq 0 \text{ is a constant}$$

Want to solve $\min_{x \in S^{n-1}} f(x)$

$$\frac{\partial}{\partial x} f(x) = A x + \beta x^3 \qquad \begin{pmatrix} x_1^3 \\ \vdots \\ x_n^3 \end{pmatrix} \qquad \frac{\partial^2}{\partial x^2} f(x) = A + 3\beta \begin{pmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{pmatrix}$$

- $\bar{M} = \mathbb{R}^n$, $\bar{f} : \bar{M} \longrightarrow \mathbb{R}$

$$x \longmapsto \frac{1}{2} x^T A x + \frac{\beta}{4} \sum_{i=1}^{n} x_i^4$$

$T_x \bar{M} = \mathbb{R}^n$, pick a metric

$$\bar{g}_x(u, v) = v^T G(x) u$$

$$g_x(\operatorname{grad} \bar{f}(x), v) = \left\langle \frac{\partial}{\partial x} f(x), v \right\rangle, \ \forall v \in T_x \bar{M} = \mathbb{R}^n$$

$$\Rightarrow \quad G(x) \operatorname{grad} \bar{f}(x) = \frac{\partial}{\partial x} f(x)$$

$$\Rightarrow \quad \operatorname{grad} \bar{f}(x) = G(x)^{-1} [A x + \beta x^3]$$

- $M = S^{n-1}$, consider the embedded Riemannian submanifold (M, g) in $(\bar{M}, \bar{g})$

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

$$x \longmapsto \frac{1}{2} x^T A x + \frac{\beta}{4} \sum_{i=1}^{n} x_i^4$$

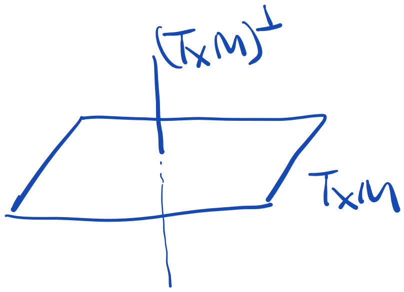
$$g_x(\operatorname{grad} f(x), v) = \bar{g}_x(\operatorname{grad} \bar{f}(x), v), \quad \forall v \in T_x M$$

To find $\operatorname{grad} f(x) \in T_x M \subseteq T_x \bar{M} = \mathbb{R}^n$,

we regard $\operatorname{grad} f(x)$, v as elements in $T_x \bar{M}$

$$\Rightarrow \bar{g}_x(\operatorname{grad} f(x), v) = \bar{g}_x(\operatorname{grad} \bar{f}(x), v), \quad \forall v \in T_x M$$

$$\Rightarrow \bar{g}_x(\operatorname{grad} f(x) - \operatorname{grad} \bar{f}(x), v) = 0, \quad \forall v \in T_x M$$



$$T_x \bar{M} = T_x M \oplus (T_x M)^\perp$$

$\Rightarrow \operatorname{grad} f(x) - \operatorname{grad} \bar{f}(x) \perp T_x M$ in $T_x \bar{M}$

$\Rightarrow \operatorname{grad} f(x)$ is "Proj" of $\operatorname{grad} \bar{f}(x)$ on $T_x M$

$\quad\to$ in the sense of $\perp$ using $\bar{g}_x$

$M = S^{n-1}$

① $T_x M = \{ v \in \mathbb{R}^n : x^T v = 0 \} \to$ independent of g

$w \in \underbrace{(T_x M)^\perp}_{\text{depends on } \bar{g}}$, $\forall v \in T_x M$, $\bar{g}_x(w, v) = 0$

$$v^T G(x) w = 0$$

$$\Rightarrow G(x) w \parallel x$$

$$\Rightarrow (T_x M)^\perp = \{ a G(x)^{-1} x, \ a \in \mathbb{R} \}$$

② "Proj" of $\operatorname{grad} \bar{f}(x)$ onto $(T_x M)^\perp$ is

$$\frac{g_x(\operatorname{grad} \bar{f}(x), G(x)^{-1} x)}{g_x(G(x)^{-1} x, G(x)^{-1} x)} G(x)^{-1} x$$

③ "Proj" of $\operatorname{grad}\bar{f}(x)$ onto T_xM is

$$\Rightarrow \operatorname{grad} f(x) = \operatorname{grad}\bar{f}(x) - \frac{g_x(\operatorname{grad}\bar{f}(x), G(x)^{-1}x)}{g_x(G(x)^{-1}x, G(x)^{-1}x)} G(x)^{-1}x$$

$$= G(x)^{-1}[Ax + \beta x^3] - \frac{\langle \operatorname{grad}\bar{f}(x), G(x) G(x)^{-1}x \rangle}{\langle G(x)^{-1}x, G(x) G(x)^{-1}x \rangle} G(x)^{-1}x$$

$$= G(x)^{-1}[Ax+\beta x^3] - \frac{\langle \operatorname{grad}\bar{f}(x), x \rangle}{\langle G(x)^{-1}x, x \rangle} G(x)^{-1}x$$

Choices of $G(x)$:

① $G(x) = I$ meaning S^{n-1} with $g = \langle , \rangle$

② $G(x) = A$

$$\operatorname{grad}\bar{f}(x) = x + \beta A^{-1}x^3$$

③ $G(x) = A + \beta \begin{pmatrix} x_1^2 & & & \\ & x_2^2 & & \\ & & \ddots & \\ & & & x_n^2 \end{pmatrix}$

$$\operatorname{grad}\bar{f}(x) = x \Rightarrow \operatorname{graf}(x) = x - \frac{\langle x, x \rangle}{\langle G(x)^{-1}x, x \rangle} G(x)^{-1}x$$

A numerical example: $A = -\Delta + V$

Remark: an alternative way to compute grad f

$$\bar{g}_x(\operatorname{grad} f(x) - \operatorname{grad}\bar{f}(x), v) = 0, \ \forall v \in T_xM$$

$$v^T G(x) \operatorname{grad} f(x) = v^T G(x) \operatorname{grad}\bar{f}(x), \ \forall v \in T_xM$$

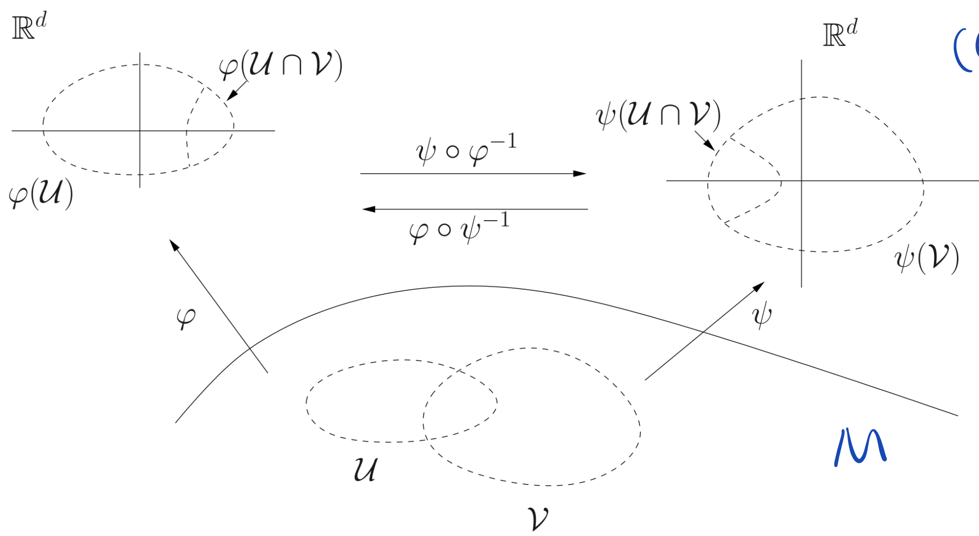
$$\Rightarrow \quad v^T [G(x) \operatorname{grad} f(x) - G(x) \operatorname{grad}\bar{f}(x)] = 0$$

$\Rightarrow$ $G(x) \operatorname{grad} f(x)$ is Proj of $G(x) \operatorname{grad}\bar{f}(x)$ onto T_xM

$$\Rightarrow G(x)\,\mathrm{grad}\, f(x) = P_{T_xM}\left[G(x)\,\mathrm{grad}\, \bar{f}(x) \right]$$

Summary of some Concepts for Riemannian Opt

- Manifold, charts, smooth manifold

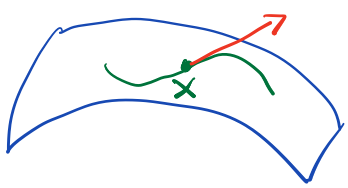


(φ, U) and (ψ, V) are coordinate charts

- Tangent Space T_xM

The general definition of tangent space of a manifold $M \subseteq \mathcal{E}$

$$T_xM = \{ c'(0) \mid c: (-\varepsilon, \varepsilon) \to M \text{ is smooth } \& \ c(0) = x \}$$



$c(t)$ is a curve on M

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

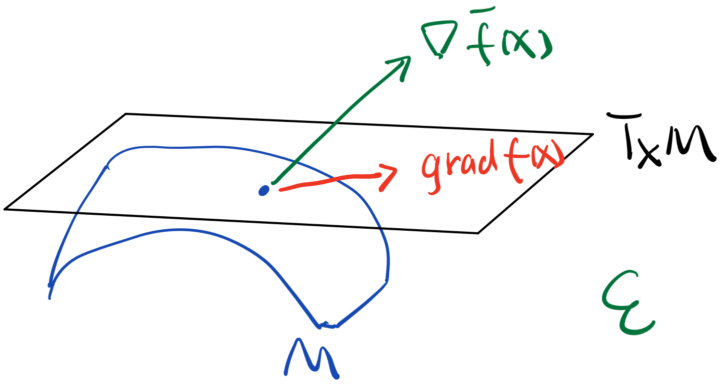
If $M \subseteq \mathcal{E}$, T_xM is a subspace in $\mathcal{E}$.

- Riemannian Metric g

at $x \in M$, $g_x(u, v)$ is an inner product for T_xM

T_xM stays the same for different choices of g

- Riemannian Gradient $\operatorname{grad} f(x) \in T_xM$



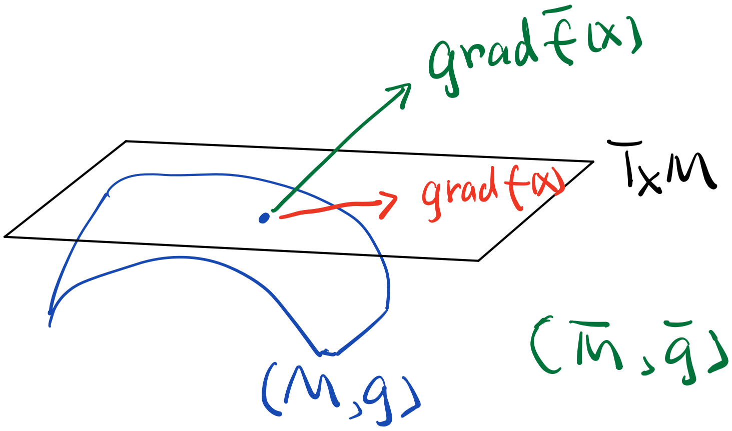
$M \subset \mathcal{E}$

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

$\bar{f}: \mathcal{E} \to \mathbb{R}$

$g_x(\operatorname{grad} f(x), v) = \langle \nabla \bar{f}, v \rangle$

$\forall v \in T_xM$



$M \subset \mathcal{E}$

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

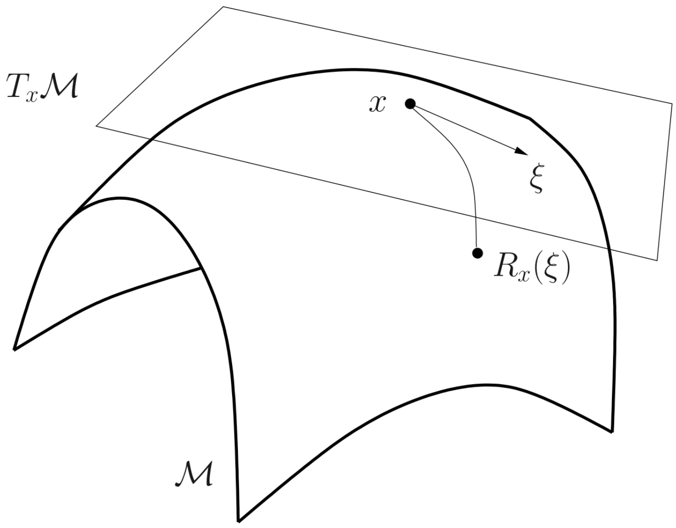
$\bar{f}: \bar{M} \to \mathbb{R}$

$g_x(\operatorname{grad} f(x), v) = \bar{g}_x(\operatorname{grad} \bar{f}(x), v)$

$\forall v \in T_xM$

	Manifold (S^{n-1})	Embedding space ($\mathbb{R}^n$)
cost	$f(x) = x^T Ax,\ x \in S^{n-1}$	$\overline{f}(x) = x^T Ax,\ x \in \mathbb{R}^n$
metric	induced metric	$\overline{g}(\xi, \zeta) = \xi^T \zeta$
tangent space	$\xi \in \mathbb{R}^n : x^T \xi = 0$	$\mathbb{R}^n$
normal space	$\xi \in \mathbb{R}^n : \xi = \alpha x$	$\emptyset$
projection onto tangent space	$\mathrm{P}_x \xi = (I - xx^T)\xi$	identity
gradient	$\operatorname{grad} f(x) = \mathrm{P}_x \operatorname{grad} \overline{f}(x)$	$\operatorname{grad} \overline{f}(x) = 2Ax$
retraction	$R_x(\xi) = \mathrm{qf}(x + \xi)$	$R_x(\xi) = x + \xi$

- Retraction at x is a map $R_x : T_xM \to M$

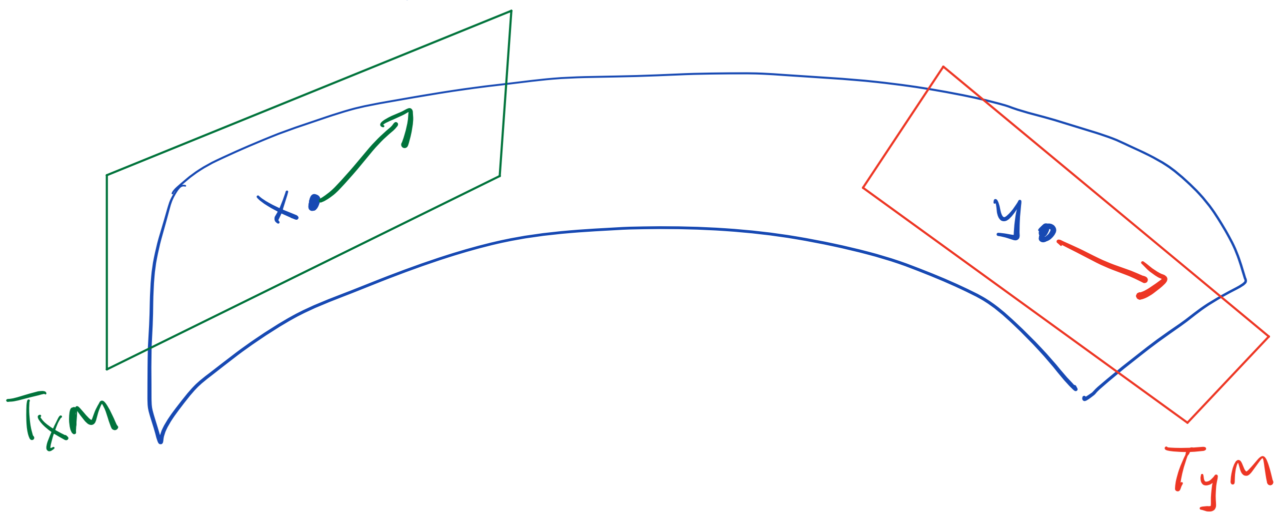


- Riemannian Gradient Descent on (M, g)

$$x_{k+1} = R_{x_k}(-\eta \,\mathrm{grad}\, f(x_k))$$

Concepts that we do not cover but are important:

- Vector transports (how to define $\nabla f(x_k) + \nabla f(x_{k-1})$)



- Riemannian Hessian

$$\mathrm{Hess} f(x) : T_xM \to T_xM$$
$$v \longmapsto \nabla_v \,\mathrm{grad} f(x)$$

∇_v is called connection

↪ covariant derivative.

- Quotient Manifold

  Grassmann Manifold

  $\mathrm{Grass}(p,n) = \{$ subspaces of dim p in $\mathbb{R}^n \}$